Solids on Soli: Millimetre-Wave Radar Sensing through Materials

ANONYMOUS AUTHOR(S)
SUBMISSION ID: PAPERS_600

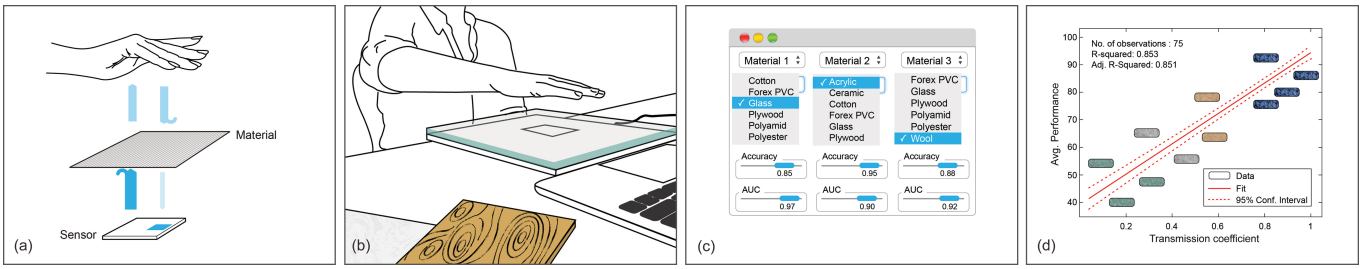

Fig. 1. Integrating radar-based gesture recognition in consumer devices requires the signal to pass through a covering material twice (a). Therefore, application designers need to carefully select suitable materials. For this, we provide an ample catalogue of 75 everyday materials that is agnostic to the underlying gesture classifier. Designers need to evaluate their own recognisers against just 3 reference materials (b), enter the observed performance measures into a simple tool (c), and get updated performance estimates for all materials in the catalogue (d).

Gesture recognition with miniaturised radar sensors has received increasing attention as a novel interaction medium. The practical use of radar technology, however, often requires sensing through materials. Yet, it is still not well understood how the internal structure of materials impacts recognition performance. To tackle this challenge, we collected a large dataset of 12,120 radar recordings for 6 paradigmatic gesture classes sensed through a variety of everyday materials, performed by a human (6 materials) and a robot (75 materials). Next, we developed a hybrid CNN+LSTM deep learning model and derived a robust method to measure signal distortions, which we used to compile a comprehensive catalogue of materials. Our experiments show our method can estimate how different materials would affect gesture recognition performance by considering just 3 reference materials from the catalogue. Our software, models, data collection platform, and labeled datasets are publicly available.

CCS Concepts: • **Hardware** → *Sensor devices and platforms*; • **Human-centered computing** → **Gestural input**; *Interaction design process and methods.*

Additional Key Words and Phrases: Radar Interaction; Materials; Soli; Gestures; Deep Learning

ACM Reference Format:
Anonymous Author(s). 2018. Solids on Soli: Millimetre-Wave Radar Sensing through Materials. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/1122445.1122456

Woodstock '18, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

In recent years, gesture recognition with miniaturised radar sensing has received increased attention in academia and industry. Two factors, in our opinion, have fuelled this trend: the emergence of low-cost low-power radar chips and the impressive breakthroughs in deep learning, which have made it possible to interpret radar signals very accurately for interaction. However, there are still several unresolved research problems in order to enable practical applications of this technology.

First and foremost, real-world interaction with radar devices often requires sensing through materials. This not only includes interaction e.g. with a mobile phone while in our pocket but also interaction with other devices (e.g. infotainment systems or smart homes) through a radar sensor that is integrated in various objects on and around us, such as clothing accessories (e.g. ties, buttons, and jewellery), car dashboards, seats, doors, indoor furniture, and walls. Nevertheless, it is still not well understood how different materials affect gesture recognition performance. Ideally, application designers would speed up development with a convenient method for determining which materials would perform best for a recogniser of their choice (Figure 2).

To make radar-based gesture recognition practical on different materials, we first need to understand how the radar signal degrades when passing through them. This seemingly straightforward problem is in reality rather challenging. One would need to acquire an expensive vector network analyser and operate it in a shielded environment, preventing electromagnetic interference with measurements. The next problem arises from the plethora of potential materials and various thicknesses one would like to evaluate. Together with different gesture sets and custom classifiers, this would result in a vast number of measurements to watch for. Lastly, if we were able to overcome all these aforementioned problems, how could we conveniently predict gesture recognition performance on untested materials when only signal characteristics (e.g. transmission coefficient) for the materials are known?

Fig. 2. Overview of materials in our catalogue as a Sankey diagram [Schmidt 2008]. Our catalogue covers 48 types of materials with varying thicknesses, 75 materials in total.

To address these challenges we (i) used the Google Soli millimetre-wave (mm-wave) radar sensor with its accompanying software development kit (SDK) to record range Doppler images and time series data (12,120 gesture executions) for 11 core features as well as 9 meta features; (ii) developed a state-of-the-art hybrid CNN+LSTM deep learning model to test recognition performance on 75 materials; and (iii) derived a robust method to measure signal distortions, which we used to compile a catalogue of materials, and predict recognition performance on arbitrary gesture classifiers.

Through visualisation of the time series data of incident and transmitted signal, we observed a strong inverse correlation of amplitude and material thickness (the greater the thickness, the smaller the amplitude) for several signal's core features that we describe later. We validate our findings with correlation analyses on 7 different material types; e.g., *Paper* and *Drywall* exhibit a strong inverse correlation and more transparent materials like *Styrofoam* exhibit weak correlations. This shows that our proposed indirect measurement method is adequate and implies that some materials are more apt for interaction than others. Based on our findings, we have compiled an extensive catalogue of everyday materials for radar sensing that we make publicly available (see Supplementary Material).

We further validate our methodology by predicting gesture recognition performance on arbitrary classifiers, extrapolating performance metrics from our catalogue. We show that it is possible to estimate gesture recognition performance for all materials in the catalogue when performance for just 3 reference materials is known.

In sum, we make the following contributions:

- A robust methodology for systematic characterisation of material properties for radar-based gesture interaction that does not involve expensive equipment and that allows to recalibrate performance estimates for the task at hand.
- A catalogue of 75 everyday materials that can be used by interaction designers to select suitable materials through which gestures can be sensed with high accuracy, regardless of the gesture set and underlying recogniser.
- Models, software, data collection platform, and datasets. This will facilitate other researchers to replicate our results and expand our catalogue, as well as develop new methods of performance estimation.

2 BACKGROUND AND PROBLEM DEFINITION

Radar technology uses radio-frequency (RF) electromagnetic waves to detect nearby objects. Essentially, a radar device has a transmitter antenna (Tx) that emits an RF pulse to the environment, and a receiving antenna (Rx) that captures the echoed pulse. The received signal is analysed to determine object properties, such as their radar cross-section [Skolnik et al. 1980, page 33] and their movement [Skolnik et al. 1980, page 101]. RF waves can be affected by a variety of phenomena such as: absorption, refraction, diffraction, polarisation, scattering, and reflection [Richard Rudd , Ken Craig , Martin Ganley 2014]. It is very difficult to analyse the effect of each of these phenomena individually, as they happen simultaneously and interact with one another. Instead, a simplified model based on the lightwave analogy is used (see Figure 3), where only the incident, reflected, and transmitted signal are considered. It is typically employed in network analysis, where designers and manufactures of networking components characterise how such components distort the input signal over a desired frequency range [Agilent 1997].

When radar sensing is used for interaction though materials, the RF signal needs to pass trough any material (that is in the way) twice. Figure 3(c) illustrates this phenomena. Since different materials have different characteristics, which depend on their thickness and their dielectric properties, these characteristics will significantly affect how the signal propagates through them. This opens up the following research questions:

(1) How to accurately characterise radar signal distortion when the signal passes through different materials?
(2) How would such distortions affect gesture recognition performance?
(3) How can we conveniently predict performance of arbitrary gestures when only the characteristics of signal distortions for a given material are known?

To address the above research questions, we proceeded as follows.

(1) **Gesture recognition experiments:** To analyse radar sensing performance through different materials, we first build and evaluate a gesture recogniser that is accurate under a control condition (with no material).
(2) **Modeling signal distortion:** To characterise radar signal distortion caused by occluding materials, we find strong predictors of material properties and thicknesses.
(3) **Modeling materials performance:** To understand which material properties affect gesture recognition performance, we carry out an extensive correlation analysis and compile a catalogue of materials with several informative parameters.
(4) **Performance on arbitrary classifiers:** To demonstrate how our catalogue of materials can be used in practice, we: (i) fit a linear regression model on a small data partition where 3 reference materials are carefully chosen; and (ii) evaluate the model on the remaining materials.

3 RELATED WORK

We analyse previous research according to our main areas of interest, namely gesture interaction, sensing technologies, mm-wave radar sensors (including Google Soli), and sensing through materials.

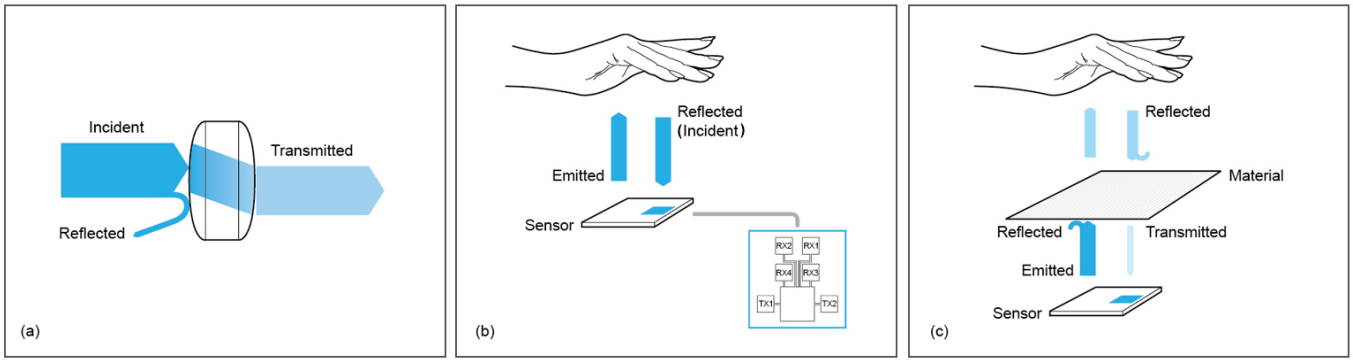

Fig. 3. Lightwave analogy (a) to characterise signal distortions [Agilent 1997]. When no occluding material is present (b), a Tx antenna sends out an RF pulse that gets reflected by the user's hand and is then received by one or more Rx antennas (referred to as *incident signal* in our proposed method to estimate radar signal distortions). The process is the same when an occluding material is present (c), however this time the reflected signal from the hand needs to pass though the material (referred to as *transmitted signal* in our proposed method).

3.1 Gesture interaction

Gesture interaction is an active research topic with a history dating back to the 1960s with Sutherland's Sketchpad project [Sutherland 1964] and a far reaching vision of the Ultimate Display essay [Sutherland 1965]. Today, gesture interaction and interfaces can be categorised into 2 broad groups: (i) mid-air gesture interaction, used for example in consumer electronics such as gaming consoles, and (ii) stroke gesture interaction, used for example in devices with touchscreens such as smartphones. We focus on the former group, given the increasing importance it has gained over the recent years.

Mid-air gesture interaction has been extensively researched as an alternative to other modes or as a complementary mode of interaction in a variety of settings such as entertainment [Bleiweiss et al. 2010; Rautaray and Agrawal 2011; Roccetti et al. 2012], automotive industry [Large et al. 2019; Molchanov et al. 2015; Pickering et al. 2007; Riener et al. 2013; Shakeri et al. 2018], medical applications [Gallo et al. 2011; Mewes et al. 2017; O'Hara et al. 2014; Rahman et al. 2015; Starner et al. 2000], wearable computing [Arefin Shimon et al. 2016; Kellogg et al. 2014; Kim et al. 2007; Kratz and Rohs 2009], smart home control [Kühnel et al. 2011; Wan et al. 2014; Wu and Wang 2012], virtual reality manipulation [Lu et al. 2012; Yang et al. 2019], and art installations [Paradiso et al. 1997; Paradiso 1999]. Such interaction is particularly interesting where other modes are dangerous, hard, or impossible to use.

3.2 Sensing technologies

Common technologies used for implementing gesture recognisers include RGB cameras [Mistry and Maes 2009; Song et al. 2014; Van Vlaenderen et al. 2015], infrared cameras [Galna et al. 2014; Kim et al. 2007; Li 2012; Song et al. 2012; Starner et al. 2000] (both also in combination with markers), ultrasound gesture recognition [McIntosh et al. 2017b; Mujibiya et al. 2013; Shakeri et al. 2018], various sensors such as electromyogram sensors, flex or bend sensors, Hall-effect sensor, and 3D accelerometers [Roccetti et al. 2012; Zhang et al. 2009], magnetic and electric field sensing [Chen et al. 2013; Cohn et al. 2012; Hwang et al. 2013; Le Goc et al. 2014], and RF-based solutions including radar [Nasr et al. 2016; PourMousavi et al. 2013], Wi-Fi [Adib et al. 2014; Pu et al. 2013; Zhao et al. 2018], GSM [Zhao et al. 2014], and RFID [Kellogg et al. 2014].

RF solutions have several advantages over other technologies. Above all, they are insensitive to light, which usually affects camera and especially IR based solutions (both cannot be used in bright sunlight). They do not require an elaborate setup of various sensors on or around users. In addition, the RF signal can penetrate non-metallic occluding material(s) and can sense objects and their movements through them.

RF sensing has been used for analysing walking patterns or gait, body movement signatures, and kinematics for people recognition [Chen et al. 2006; Otero 2005; Wang and Fathy 2011], tracking sleep quality and breathing patterns [Rahman et al. 2015; Zhuang et al. 2015], and recognising movements of body parts such as hands for interactive purposes [Kellogg et al. 2014; McIntosh et al. 2017a; Molchanov et al. 2015; Paradiso et al. 1997; Paradiso 1999; Pu et al. 2013; Wan et al. 2014]. The radars used in these studies operated at various frequencies, ranging from 2.4 GHz [Wan et al. 2014; Zhuang et al. 2015] to 24 GHz [Molchanov et al. 2015; Rahman et al. 2015].

3.3 Millimeter-wave radar-on-chip sensors

To detect and recognise fine-grained finger motions and dynamic complex hand gestures, it is necessary to increase the radar spatial resolution. For this, radar chips working at even higher frequencies, around 50–70 GHz, have been recently used [Nasr et al. 2016; Wei and Zhang 2015]. Such chips open up the path to precise close-range gesture interactions in a variety of applications including wearable, mobile, and ubiquitous computing.

The mm-wave radar sensors offer several advantages. Since these sensors are operating in the mm-wave range, they also allow a much higher integration of the circuit due to the reduced geometry of different passive (non-moving) components and require low-power to work [Nasr et al. 2016]. These properties also allow manufacturing the chip inexpensively at scale, as well as easy integration into a variety of devices. A mm-wave radar sensor can also recognise gestures in 3D and overlapping fingers [Lien et al. 2016].

The research published so far has demonstrated that mm-wave high-frequency radar sensing is very effective in detecting close-proximity, subtle, nonrigid motion micro-gestures, mostly articulated with hands and fingers (e.g. rubbing, pinching, or swiping) [Lien et al. 2016; Wang et al. 2016] or with small objects (e.g. pens) [Wei and Zhang 2015]. Recent research includes, among others, exploration of interaction possibilities combining micro-gestures with large hand gestures [Ens et al. 2017], providing interaction capabilities to everyday non-interactive objects through wrist wearing mm-wave radar and augmented reality [Čopič Pucihar et al. 2019], as well as creating and manipulating music [Bernardo et al. 2017; Sandor and Nakamura 2018]. A mm-wave radar sensor can also distinguish various materials when placed on top of it [McIntosh et al. 2017a; Yeo et al. 2016]. What is missing, however, is an investigation of gesture recognition performance *through* various materials present on and around us, which is the focus of our work.

3.4 Sensing through materials

The fact that RF signals can penetrate non-metallic materials makes them particularly interesting for interaction, as sensors can be embedded in everyday objects and surfaces. Alas, sensing through materials has been barely explored. Notable examples include tracking people through walls [Adib et al. 2015; Chetty et al. 2011] and gesture recognition through walls [Pu et al. 2013] and above an office desk [McIntosh et al. 2017a]. All these approaches have focused on coarse gestures instead of fine-grained ones, have considered only one material, and did not use radar-on-chip sensors.

Leiva et al. [2020] investigated radar-based mid-air gesture recognition on wearable devices, but they did not characterise radar signal distortions nor estimated recognition performance on arbitrary gesture classifiers. Now that mm-wave radar technology is available on consumer products, it is expected that it will be further integrated in a variety of objects on and around us. Therefore, our work is both timely and important, as it will help others select suitable materials for their applications.

3.5 Google Soli

To study fine-grained micro-gestures and sensing through materials, we used Google Soli – a 60 GHz 4-channel receiver (Rx antennas) 2-channel transmitter (Tx antennas) radar-on-chip. Complemented with time-varying micro-Doppler frequency features analysis [Chen et al. 2006], the sensor offers detection of movements with near-millimeter accuracy [Nasr et al. 2016].

Soli comes with an SDK that is capable of representing the radar signal with range Doppler images as well as a variety of extracted low-level features, also called *core* features [Lien et al. 2016]:

(F1) *Range:* Overall distance of the moving targets to the sensor.
(F2) *Acceleration:* Overall acceleration of the moving targets.
(F3) *Energy total:* Amount of reflected energy overall.
(F4) *Energy moving:* Amount of reflected energy from the moving targets.
(F5) *Velocity:* Overall velocity of the moving targets.
(F6) *Velocity dispersion:* Dispersion of energy over the Doppler space.
(F7) *Spatial dispersion:* Dispersion of energy over the range space.
(F8) *Energy strongest component:* Amount of reflected energy from the most dominant moving target.
(F9) *Movement index:* Moving target identifier.
(F10) *Fine displacement:* Instantaneous velocity of each moving target.
(F11) *Velocity centroid:* Weighted average of the overall velocity.

These core features essentially characterise the energy distribution across the radar signal transformation space, which have been shown to accurately describe relative finger dynamics [Bernardo et al. 2017; Lien et al. 2016] as well as end effector trajectories [Molchanov et al. 2015; Zhang and Cao 2018].

The Soli sensor has become readily available in consumer electronics. It is for example integrated in the Google Pixel 4, to control some of the phone's functions such as swiping in the music player to play the previous or next song, as well as for detecting approaching faces. While the Soli sensor has not been included in the Google Pixel 5, the technology is being further developed and it is expected to be soon included in smart speakers and smartwatches.

4 GESTURE RECOGNITION EXPERIMENTS

As hinted previously, we developed a gesture recogniser that is accurate under a control condition (with no material). This recogniser will be our baseline classifier to analyse radar sensing performance through different materials.

4.1 Gesture set

To decide on a suitable set of gestures for investigating our research questions, we surveyed previous work that used mm-wave radar sensing [Wang et al. 2016] as well as gestures supported by the Soli sensor in consumer devices. From these gesture sets, we chose the ones that are relatively easy for a mechanical system to replicate. Eventually, a set of 6 distinct gestures was selected (Figure 4).

4.2 Experimental system

To explore the effect of different materials on gesture classification performance, we should be able to execute as many times as needed any number of gestures. For this reason, in addition to recruiting a human participant, we built a mechanical system capable of repeating the same movement above a radar sensor.

4.2.1 Robot system. Our system follows the setup proposed by Leiva et al. [2020]. It consists of (i) a robot-based gesture simulator based on the GoPiGo3, to which an empty plastic ball of 5 cm diameter is attached; (ii) an aluminium-shielded frame placed on the table, on which different materials can be placed (the frame prevents the radar signal from escaping around the analysed material); and (iii) the Soli radar sensor placed in the frame and connected to a computer (see Figure 5 left). The distance between the sensor and the ball in its lowest position is always set to 5 cm.

The robot platform is designed to generate two types of movements: pendulum-like (e.g. swing or swipe movements) and vertical movements along the z-axis. A pendulum movement is generated by manually releasing the ball from a limiter position, whereas the vertical movements are automatically generated by the robot.

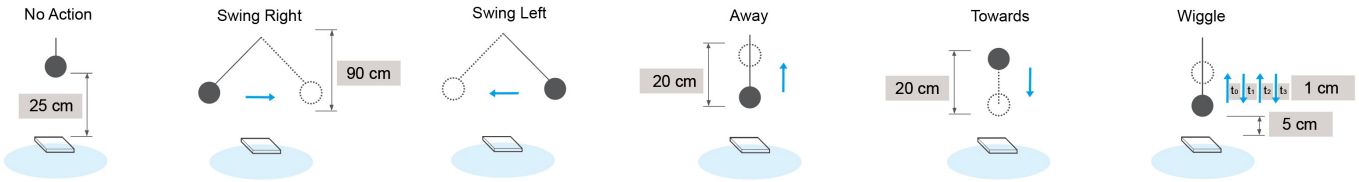

Fig. 4. Experimental gesture set. Filled and dashed circles denote, respectively, the initial and final position of the ball.

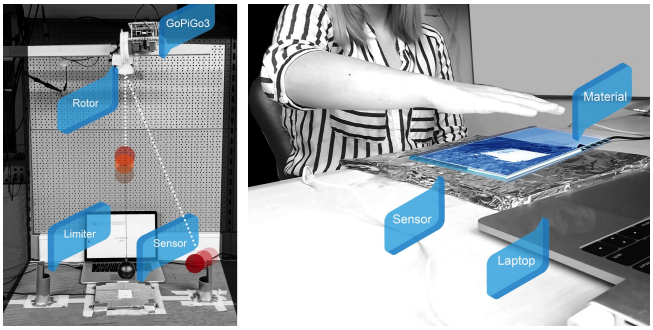

Fig. 5. Experimental platform setup. Left: robot setup. Right: human setup.

4.2.2 Human system. This setup is identical to the robot system, but instead of a ball, a human hand executes the gestures (Figure 5 right). A user performs the gestures while sitting in front of a computer display that instructs what gesture to perform and when to perform it. Besides visual indicators, the system also plays a sound in order to help the user executing the gesture systematically within a fixed time window.

4.3 Data collection

Each gesture articulation is stored in two formats: (i) a sequence of frames of range Doppler images; and (ii) a time series of 11 core features, calculated by the Soli SDK. Each frame of the range Doppler images is 32×32 px. For training and testing our model we only used range Doppler images, since they enable robust gesture recognition [Čopič Pucihar et al. 2019; Wang et al. 2016]. The extracted core features were used to characterise distortions of the signal caused by materials.

The sensor was configured to record at 1000 Hz, the Doppler range sensitivity was set to [-2,0] dB, and the built-in adaptive clutter filter was disabled. We chose 1000 Hz sampling because it represents an upper bound (it is unlikely that a higher sampling frequency would improve gesture recognition performance) which we can easily downsample as needed (note that upsampling the frame rate is a more challenging operation, as it has to "hallucinate" interpolated frames). The Doppler range sensitivity was set based on visual inspection of the captured frames as our robot system performed the different gestures.

4.3.1 Robot. We collected 20 repetitions of 6 distinct gestures (Figure 4) for each of the 75 materials that can be found in the Supplementary Material. This resulted in 9,000 gesture recordings. We also recorded 200 trials of 6 distinct gestures for no material (baseline condition), 1,200 gesture recordings in total.

4.3.2 Human. We recorded gesture data for 6 different material conditions. This resulted in 720 recordings (6 gestures x 6 materials x 20 repetitions). We also captured 200 recordings of 6 distinct gestures for the no material condition, 1,200 gesture recordings in total, which we used as the baseline condition. All recordings were performed by the same participant, aimed at illustrating the use case of building a personalised classifier [Lien et al. 2016] as well as facilitating both the data acquisition and the comparison against robot data. The participant proceeded in 2 different sessions of 600 recordings each.

4.4 Data preprocessing

Since Soli computes range Doppler images for each of the 4 Rx antennas, we averaged them as a single image to ensure a robust frame representation. Further, frame images were grayscaled and sequences were resampled to 100 Hz and padded to 400 timesteps, which is large enough to accommodate for arbitrary gesture articulations. As a reference, each of the recorded gestures took on average 1.5 seconds, or 150 timesteps.

Data resampling contributes to speed up preprocessing, training, and testing of our gesture recogniser (to be described next). We decided to resample frame sequences to 100 Hz based on pilot studies, as the performance of our recogniser only started to degrade significantly at sampling frequencies lower than 100 Hz.

4.5 Model architecture

We built a hybrid deep CNN+LSTM (convolutional neural network + long short-term memory) model, inspired by previous work on human activity recognition [Hammerla et al. 2016; Lien et al. 2016; Ng et al. 2015; Ordóñez and Roggen 2016], which is depicted in Figure 6. Our model processes each frame (Doppler image) by means of a stack of $32 \times 64 \times 128$ convolutional layers with 3×3 filters to capture spatial information. The result is processed in a recurrent fashion by means of an LSTM layer to capture temporal information, and eventually classified with a softmax layer. Our model has 1.2M weights, which is small for today's standards.

Each convolutional layer automatically extracts feature maps from the input frames that are further processed by max pooling

and spatial dropout layers. The max pooling layers (pool size of 2) downsample the feature maps by taking the largest value of the map patches, resulting in a local translation invariance.

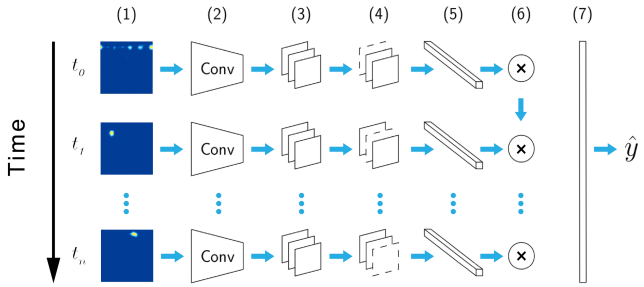

Fig. 6. Hybrid deep learning model architecture. Range Doppler images (1) are processed with a CNN (2) that extracts feature maps followed by max pooling (3) and spatial dropout (4) layers. Then, a fully connected layer (5) creates the feature vectors for a recurrent LSTM layer with dropout (6) and finally a softmax layer (7) outputs the gesture class prediction ($\hat{y}$).

Crucially, the spatial dropout layer (drop rate of 0.25) removes entire feature maps at random, instead of individual neurons (as it happens in regular dropout layers), which promotes independence between feature maps, thus improving performance. Finally, the LSTM layer (embedding size of 128) uses the computed feature maps to classify the frame sequences in a recurrent fashion, this is, the information embedded at each frame depends on the previous frame. The LSTM layer uses both a dropout rate and a recurrent dropout rate of 0.25. The softmax layer has dimensionality of 6, since we have 6 gesture classes.

4.6 Model training and evaluation

We created random splits comprising 50% of the data for model training, 20% for model validation, and the remaining 30% for model testing. The test data are held out as a separate partition, which simulates unseen data. We also used a simple data augmentation technique: each training sequence was randomly left- and right-trimmed to 10% of the original sequence length, dropping both initial and ending frames. The model was trained only with data from no material condition.

We used the Adam optimiser with learning rate $\eta = 0.0005$ and decay rates $\beta_1 = 0.9$, $\beta_2 = 0.999$. The model was trained in batches of 10 sequences each using categorical cross-entropy as loss function. The maximum number of epochs was set to 200, but we also set an early stopping criteria of 50 epochs, which means that training stopped if the validation loss did not improve after 50 consecutive epochs, and the best model weights obtained up to that moment were retained.

We built 7 classifiers for each source of data (robot and human), totaling 14 classifiers. As can be seen in Table 1, all classifiers were built for different combinations of 6 gestures recorded (6/6 and all possible combinations of 5/6 gestures). This was done in order to demonstrate that our findings are agnostic to the recogniser and data source.

We evaluated each classifier on the baseline (no material) condition as well as on 75 materials for the robot hand and on 6 materials for the human hand, respectively. For the baseline condition, as stated previously, the model was tested on a held-out partition comprising 30% of the recorded data. For the material conditions, the model was tested on all the recorded data, as none of the materials were used for training.

4.7 Results

Table 1 shows that our model architecture performs well for different combinations of gestures: accuracy is above 96% for the robot and above 87% for the human. The overall performance is better in the robot condition, as expected, since the variation in gesture articulation is much lower. For human data, the model had some difficulties in distinguishing between swipe-right and swipe-left, as well as between towards and wiggle gesture (see Table 2). Therefore, when one of these gestures is omitted from the gesture set, accuracy improves. For example for [G0 G2 G3 G4 G5] gesture set, accuracy improves to 98.67% (from 90.8%) for all classifiers; see Table 1. G0 denotes no action, which is used as a "rejection class". Further evaluation on all material conditions are provided in the Supplementary Material.

Table 1 Gesture recognition performance metrics for the baseline (no material) condition.

Source	Gesture set	Recordings	ACC	AUC	Precision	Recall	F1
Robot	[G0 G1 G2 G3 G4 G5]	1,200	97.35	98.42	97.40	97.35	97.35
Robot	[G1 G2 G3 G4 G5]	1,000	96.07	97.57	96.20	96.07	96.10
Robot	[G0 G1 G2 G3 G4]	1,000	99.29	99.55	99.31	99.29	99.29
Robot	[G0 G1 G2 G3 G5]	1,000	96.55	97.88	96.75	96.55	96.56
Robot	[G0 G1 G2 G4 G5]	1,000	98.97	99.35	98.97	98.97	98.97
Robot	[G0 G1 G3 G4 G5]	1,000	98.57	99.11	98.60	98.57	98.57
Robot	[G0 G2 G3 G4 G5]	1,000	99.29	99.55	99.29	99.29	99.29
Human	[G0 G1 G2 G3 G4 G5]	1,200	90.83	94.50	91.46	90.83	90.74
Human	[G1 G2 G3 G4 G5]	1,000	87.00	91.88	87.39	87.00	87.06
Human	[G0 G1 G2 G3 G4]	1,000	92.00	95.00	92.85	92.00	91.82
Human	[G0 G1 G2 G3 G5]	1,000	93.33	95.83	93.31	93.33	93.29
Human	[G0 G1 G2 G4 G5]	1,000	92.67	95.42	92.78	92.67	92.56
Human	[G0 G1 G3 G4 G5]	1,000	98.67	99.17	98.69	98.67	98.67
Human	[G0 G2 G3 G4 G5]	1,000	98.33	98.96	98.42	98.33	98.33

Gestures:
G0 - No action G1 - Swing left G2 - Swing right G3 - Towards G4 - Away G5 - Wiggle

Table 2 Confusion matrix for human data and two gesture sets.

Gesture classifier:
Gesture set: [G0 G1 G2 G3 G4 G5]
Dataset: Human
Material: none
Recordings: 1,200

	G0	G1	G2	G3	G4	G5
G0	60	0	0	0	0	0
G1	0	41	17	2	0	0
G2	0	5	55	0	0	0
G3	0	0	0	60	0	0
G4	0	2	0	1	57	0
G5	0	0	0	6	0	54

Gesture classifier:
Gesture set: [G0 G2 G3 G4 G5]
Dataset: Human
Material: none
Recordings: 1,000

	G0	G2	G3	G4	G5
G0	60	0	0	0	0
G2	1	59	0	0	0
G3	0	0	60		0
G4	0	0	0	60	0
G5	0	0	4	0	56

Gestures
G0 - No action
G1 - Swing left
G2 - Swing right
G3 - Towards
G4 - Away
G5 - Wiggle

5 MODELING SIGNAL DISTORTION

In order to characterise distortion of electromagnetic waves as they pass though materials, we need to measure differences between incident and transmitted signal across the frequency range of the radar-on-chip sensor; see Figure 3(a). As discussed in section 1, one approach to accomplish this is by means of a vector network analyser. This device can measure distortions in the signal amplitude and phase as the material is placed between the Tx and Rx antennas. If the material also inflicts a non-liner behaviour on the signal, frequency shifts as well as harmonic frequencies may be observed [Agilent 1997]. Such measurements need to be performed in a specially prepared isolated experimental environment preventing electromagnetic interference. A dedicated and expensive equipment makes this approach considerably costly.

As an alternative, we propose to perform indirect measurements of the difference between incident and transmitted signal using only information from the radar sensor. Using the mechanical system described in subsection 4.1, we can capture the radar signal as the ball swings like a pendulum above the sensor. By comparing the incident signal (the signal captured when no material is present) to transmitted signal (the signal captured when different materials occlude the sensor) we can measure distortions in the Amplitude and Signal-to-Noise Ratio, see Figure 3(b–c). We formulated the following hypotheses:

- H1: Amplitude distortions of the radar signal should be detectable using indirect measurements of core features where the distortion is predominantly linear.
- H2: Distortions of SNR should be detectable using indirect measurements also available through the above-mentioned core features.

5.1 Data collection

We captured data of the radar signal passing through 7 different materials of various thicknesses (totaling 26 materials) as our gesture simulator performed 2 pendulum swings of a ball over the sensor (see subsection 4.1 for details). We chose this gesture as it induces a signal that is sinusoidal in nature with a clear amplitude, period, and phase, which should be easy to notice visually. To make our observations robust to measurements error, we repeated the recording 20 times for each material (26 x 20 = 520 recordings in total).

5.2 Data analysis

We use 7 different materials of various thicknesses, namely: *Oriented strand board*, *Paper*, *Drywall*, *Acrylic*, *Polycarbonate*, *Polyethylene*, and *Styrofoam*.

5.2.1 *Detecting amplitude distortions with indirect measures (H1).*

Visualisation of time series data. For each material, we expect that a drop in signal amplitude will correlate with the thickness and properties of the material (e.g. the greater the thickness, the smaller the amplitude). We also expect to observe periodic behaviour of the extracted features (e.g. distinct peaks at a constant period).

Since we completed 20 recordings of the pendulum swing gesture for each material, we aggregated the data prior to visualisation. Then, we synchronised the first peak or valley on the same time series. In sum, our aggregation procedure is the following:

(1) Downsample the recording frame rate from 1000 Hz to 200 Hz, which is more than enough to capture fine-grained variations of Soli core features.
(2) Perform Dynamic Time Warping with barycenter averaging [Petitjean et al. 2011].
(3) Smooth the time series by calculating the mean over a sliding window of 10 frames.
(4) Apply Savitzky-Golay finite impulse response (FIR) filter with polynomial order 3 and frame length of 21. These hyperparameters are recommended by Schafer [2011].

We opted for a barycentric averaging method because the arithmetic mean depends on the order of frame aggregation, which can be problematic when tying to create reliable descriptors [Petitjean et al. 2011]. Further, we selected a relatively small window size (10 frames) and small polynomial order (3) to avoid aggressive filtering.

We note that core feature F10 (Fine Displacement) requires additional preprocessing to compensate for measurement drifts caused by the sensor, therefore we performed 2 additional steps:

(1) Outlier removal using a Hampel filter with a window size of 100 frames and removal criteria of 3 standard deviations.
(2) Offset the signal to zero using the minimum value in the whole time series.

Correlation analysis of amplitude and material thickness. To reinforce the visual observation of amplitude and material thickness, we ran a correlation analysis between peak-to-peak amplitude of time series data between core features and the thickness of a given material. This was done for the same 7 materials depicted before.

5.2.2 *Detecting distortions in SNR (H2).*

Signal-to-Noise Ratio is an important quality measure in digital signal processing. It is defined as the ratio of the power of a signal to the power of background noise. Similar to signal amplitude, we expect to observe a distortion in the domain of SNR when the signal passes through materials. We also expect to see an inverse correlation betwewn SNR and material thickness (the greater the thickness, the smaller the SNR).

We performed an indirect computation of SNR, as follows. First, instead of considering signal power we considered signal amplitude. Second, the amplitude measured in the "no action" gesture cases (Figure 4), in which the ball is stationary above the sensor, is considered background noise. We took 20 measurements from each of the above-mentioned 7 materials to get a reliable estimation, and aggregated these measurements by averaging at each frame of each time series. We also performed frame outlier removal with the interquartile range (IQR) technique to ensure consistent estimates.

5.3 Results

The visualisation of time series data for *Paper* materials (10, 100, and 200 sheets of paper) is provided in Figure 7. We can see that peak-to-peak amplitude drops with increased material thickness for core features *Energy total* (F3), *Energy moving* (F4), and *Energy strongest component* (F8). These features follows the periodic behaviour of a pendulum swing. On the other hand, the visualisation of time series data for more transparent materials, such as (*Styrofoam*, *Polyethylene*,

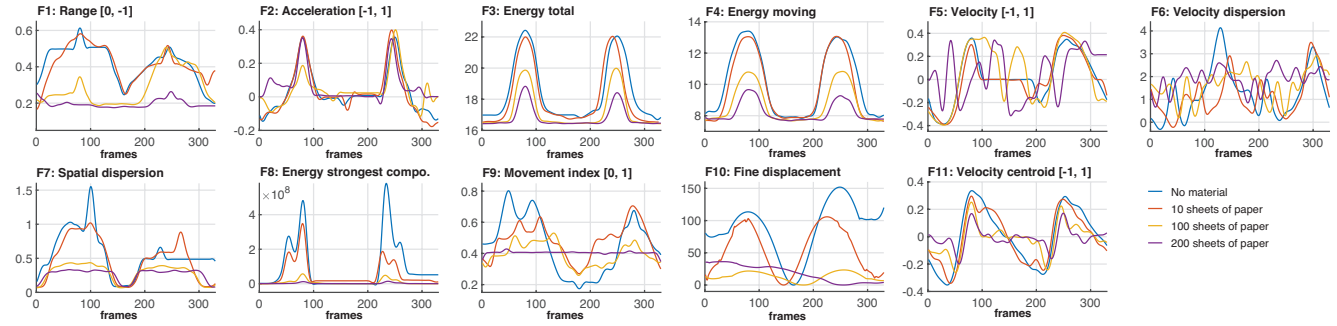

Fig. 7. Time series of the radar signal as the ball swings past the sensor. Each plot represents one Soli core feature for the baseline (no material) condition and 3 different thicknesses of *Paper* material (10, 100, and 200 sheets of paper). See our Supplementary Material for additional results.

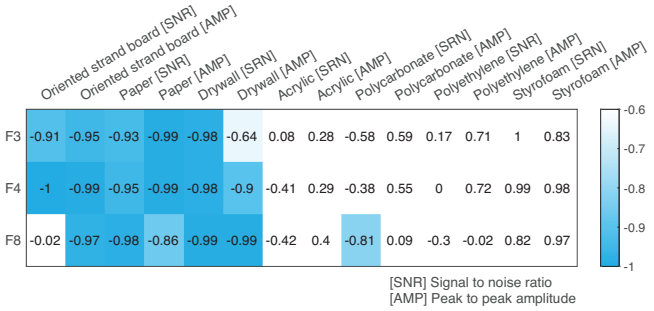

Fig. 8. Correlation between *peak-to-peak amplitude* and *Signal-to-Noise Ratio* against material *thickness*.

Acrylic, and *Polycarbonate*), only follow periodic swing behaviour whereas a drop in amplitude was not observed (see Supplementary Material).

The correlation analysis of Amplitude and SNR with material thickness (Figure 8) confirms our visual observations. Except for more transparent materials to the radar signal the inverse correlation is strong (mean -0.92 for amplitude and mean -0.82 for SNR). We can conclude that good material candidates for linking performance of gesture classifiers and signal distortions are likely to come from Soli core features F3, F4 and F8.

6 MODELING MATERIALS PERFORMANCE

As the quality of transmitted signal decreases due to signal distortions caused by occluding materials, we expect to see a performance drop in our gesture classifiers. We hypothesise:

- H3: Distortions of amplitude and SNR caused by materials will result in a drop of recognition performance.
- H4: Performance drop can be modeled using material properties that are based on (indirect) measurement of incident and transited signal.

In addition, H4 opens up the following questions:

- What material properties should be used for estimation?
- How well does a (linear regression) model fit our data?

6.1 Material properties

We define (physical) material properties as a set of meta features that can be calculated for a time series of each core feature. We base our propositions on the popular analysis methods of time series (e.g. Median, Energy, Asum, etc.) and the terminology used in network analysis domain (Transmission coefficient, Insertion loss, SNR, Amplitude).

In subsubsection 5.2.1 and subsubsection 5.2.2 we already highlighted the importance of Amplitude and SNR for characterisation of signal distortions as well as defined how these metrics can be estimated using indirect measures. From common characteristic of signal distortion used in network analysis [Agilent 1997], we selected *Transmission coefficient* and *Insertion loss*. This selection was made on the basis of the task at hand whilst also considering the limitations of our experimentation system; e.g. our system cannot measure phase shift.

Transmission coefficient [T_c] is defined as the transmitted voltage divided by the incident voltage. If the absolute value is bigger than 1, a system is said to have gain; otherwise it has attenuation [Agilent 1997]. We make the assumption that occluding materials can only induce insertion loss, hence it is reasonable to limit their transmission coefficient to 1. As in our system we can not directly measure voltage, we use signal Amplitude to calculate T_c. Our incident signal is the signal obtained when there is no material between the sensor and the swinging ball, whereas transmitted signal is the signal obtained when the material occludes the sensor as the ball swings past it; see Figure 3 (b–c).

Insertion loss L (in dB) is a pseudo-feature calculated based on the transmission coefficient [Agilent 1997]: $L = -20 \log_{10} |T_c|$. Only values bigger than 0 are reasonable. Following standards measures in time series analysis, we chose the following descriptive statistics as meta features: Mean, Median, Maximum, Asum (absolute sum across all values in the time series), and Energy (sum of squares across the whole time series).

6.2 Modeling performance drop

We build linear regression models of performance drop, where performance is defined as Avg. Perf = (Accuracy+AUC)/2 to get a single prediction outcome. We analyse how different material meta features (for each core features) correlate best with gesture recognition

performance, as different materials are placed between the sensor and the moving ball. The procedure follows these steps:

(1) Determine which core features are good candidates for the task at hand by conducting a correlation analysis of average performance against each material's meta features, for all core features.
(2) From subsection 5.3, we already know that F3, F4, and F8 are good core features. However, there might be others which we failed to identify. These can be found by looking at their correlation coefficient. Note: We ignore core features that demonstrate linear correlation amongst each other.
(3) Based on the previous feature selection, fit a linear regression model of performance given various predictors (model coefficients) and conduct a statistical evaluation. We repeat this step for all the 14 gesture classifiers (subsection 4.6)
(4) Based on the statistical results, choose the strongest predictor (with the highest p-value) and fit a simple linear regression model using only this coefficient and conduct a statistical evaluation. We also repeat this step for all the 14 gesture classifiers.

6.3 Results

6.3.1 Correlation analysis for all core features. The correlation analysis of gesture recognition performance, with core features for different material meta features, showed that the best overall correlating core feature is Energy moving (F4). The results also indicate that Acceleration (F2), Velocity (F5), Movement index (F9), and Fine displacement (F10) are not good predictors (see Figure 9). For these core features, the absolute mean and absolute median values across all material meta features are either low or have a high standard deviation (see 3 bottom rows in Figure 9). Based on these results, we conclude that core features F1, F3, F4, F6, F7, F8 and F11 should only be used for building multi-predictor linear regression model.

The high correlation between performance, amplitude, and SNR (rows 5 and 9 in Figure 9) also allow us to confirm H3: Distortions of amplitude and SNR caused by an occluding material will result in a substantial degradation of recognition performance.

The results show that the best correlating material meta features are Transmission coefficient and Insertion loss, where the two can be further enhanced if capped (limited) to 1 and 0 respectively. This capping of values is reasonable because material conditions cannot have higher amplitude than the baseline (no material) condition. For this reason, the *Transmission coefficient limited* and *Insertion loss limited* should be considered for building multi-predictor linear regression models.

6.3.2 Linear regression models. When building the linear regression models of performance, we initially chose all high-correlation core features (F1, F3, F4, F6, F7, F8, and F11). However, it turned out that *Energy strongest component* (F8) is linearly dependent to other core features. For this reason, we removed F8 from model coefficients and built a multi-predictor linear regression model with 54 predictors (9 material properties for each selected core feature), marked as bold text in Figure 9). This linear regression model with multiple predictors (Figure 10) exhibits a very good fit; e.g. Adj. R^2 is 0.94. However, if we build a simple liner regression model using

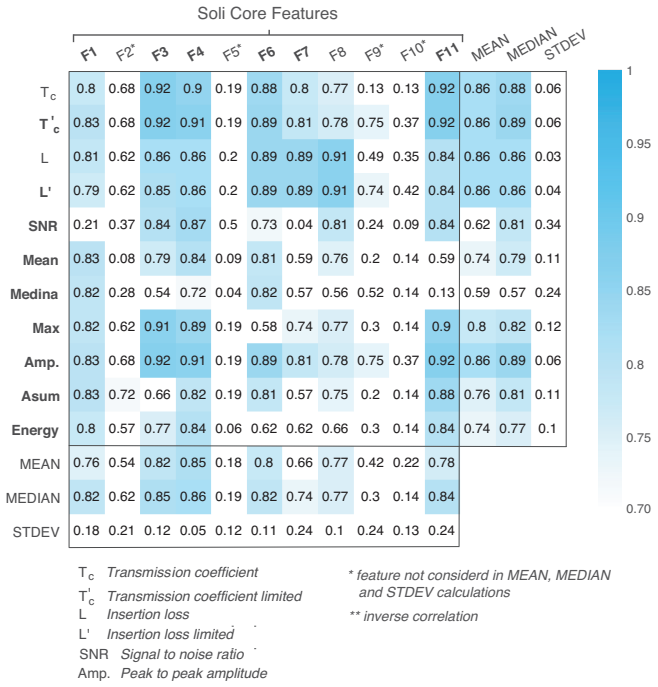

Fig. 9. Correlation analysis between gesture recognition performance and material meta features of Soli core features. Features F2, F5, F9, and F10 are weakly correlated hence they were excluded from MEAN, MEDIAN, and STDEV row-wise calculations.

only 1 predictor — e.g. the most relevant material's meta feature *Transmission coefficient limited* for core feature *Energy Total* (F3) — we get a reasonably good fit: Adj. R^2 is 0.85.

The results in Table 3 show how well linear regression models can predict performance of 14 gesture classifiers. Single and multi-predictor regression models are built in the same way as in Figure 10. The results show a good fit for all models, with a high Adj. R^2 for the multi-predictor model (mean 0.91, std 0.03) and a good Adj. R^2 for the single-predictor model (mean 0.81 with std 0.03 for robot and mean 0.78 with std 0.11 for human). This allows us to confirm H4: it is possible to build a linear regression model of gesture detection performance drop using material properties that are based on indirect measurement of incident and transited signal. This was validated for both robot and human, and for 7 gesture classifiers trained with different combinations of gestures.

7 PERFORMANCE ON ARBITRARY CLASSIFIERS

One of the goals of our catalogue (see Supplementary Material) is to conveniently predict performance of a gesture classifier when only characteristics of signal distortions for a given material are known. This would enable designers to make more informed decisions when deciding on their own gesture sets, model architectures, and operational conditions for their sensing systems.

We evaluate how well our catalogue of materials can support the above-mentioned claim for situations where performance (as

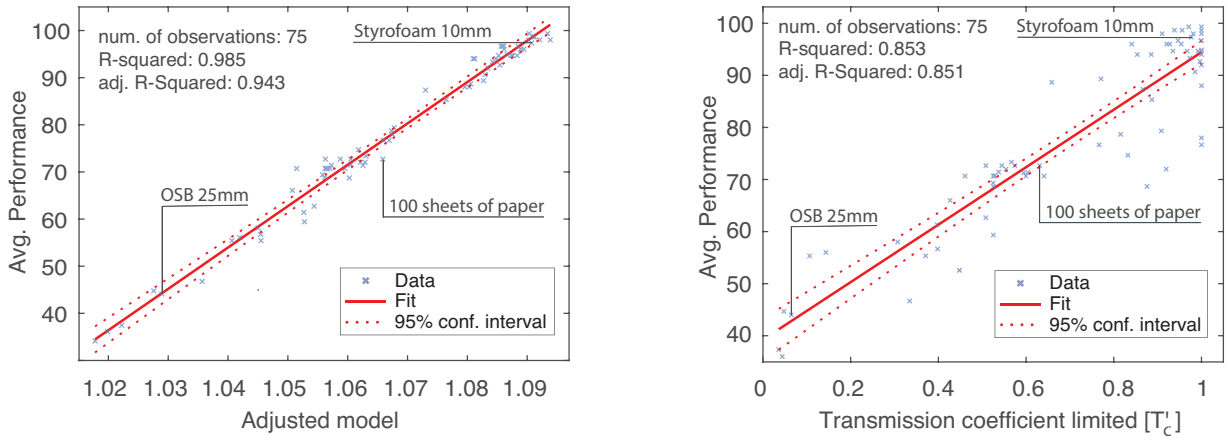

Fig. 10. Linear regression models to predict gesture classification performance. Left: multi-predictor model using 54 high-correlation Soli core features and material meta features, marked in bold in Figure 9. Right: single-predictor model using *Transmission coefficient limited* of core feature *Energy Total* (F3).

Table 3 Linear regression models (LRM) to predict performance of different gesture classifiers trained on both robot and human data with different gestures sets. The results show strong correlation across all conditions.

Source: Robot No material data only		Source: Robot, Materials: 75												Source: Human, Materials: 6					
Gesture classifiers		Single LRM Avg Perf ~ Trans. coef. limited (F3)						Multiple LRM Avg Perf ~ 54 coefficients						Single LRM Avg Perf ~ Trans. coef. limited (F3)					
Gesture set	**Recordings**	**DoF**	**RMSE**	**MAE**	**R^2**	**Adj. R^2**	**p**	**DoF**	**RMSE**	**MAE**	**R^2**	**Adj.R^2**	**p**	**DoF**	**RMSE**	**MAE**	**R^2**	**Adj.R^2**	**p**
[G0 G1 G2 G3 G4 G5]	1,200	73	6.78	5.09	0.85	0.85	<.001	20	4.20	1.64	0.99	0.94	<.001	4	10.20	6.90	0.66	0.58	<.05
[G1 G2 G3 G4 G5]	1,000	73	6.49	4.73	0.82	0.81	<.001	20	4.90	1.92	0.97	0.89	<.001	4	3.70	2.10	0.90	0.87	<.05
[G0 G1 G2 G3 G4]	1,000	73	8.63	6.70	0.83	0.83	<.001	20	4.88	1.85	0.99	0.94	<.001	4	4.86	3.09	0.87	0.83	<.05
[G0 G1 G2 G3 G5]	1,000	73	8.18	6.39	0.77	0.77	<.001	20	4.60	1.78	0.98	0.93	<.001	4	5.55	3.62	0.87	0.84	<.05
[G0 G1 G2 G4 G5]	1,000	73	9.30	7.48	0.79	0.79	<.001	20	5.74	2.21	0.98	0.92	<.001	4	5.76	3.78	0.76	0.70	*<.05*
[G0 G1 G3 G4 G5]	1,000	73	6.06	4.59	0.78	0.78	<.001	20	4.68	1.87	0.97	0.87	<.001	4	6.73	4.26	0.81	0.76	<.05
[G0 G2 G3 G4 G5]	1,000	73	6.63	4.91	0.85	0.85	<.001	20	6.05	2.39	0.97	0.87	<.001	4	5.10	3.43	0.92	0.90	*<.01*
	Mean		7.44	5.70	0.81	0.81			5.01	1.95	0.98	0.91			5.99	3.88	0.83	0.78	

Gestures:
G0 - No action G1 - Swing left G2 - Swing right G3 - Towards G4 - Away G5 - Wiggle

measured by the geometric average of Accuracy and AUC, see subsection 6.2) of a given gesture classifier is known for at least 3 materials from our catalogue. Note that 3 is the minimum number of materials to derive reasonable results, in order to cover the whole range of our catalogue. We use only one predictor (transmission coefficient) for simplicity, but we have shown that is possible to improve the model fit by considering more predictors (subsection 6.2). However, at the same time, we should note that adding more predictors implies additional effort, hence a small number of predictors is preferable.

We choose appropriate materials based on the Transmission coefficient limited (T'_c) as follows: one material must come from the low opacity range ($T'_c \in [0.8, 1]$), one from the middle opacity range ($T'_c \in [0.3, 0.5]$), and one from the high opacity range ($T'_c \in [0, 0.2]$). We fit the linear regression model with this single coefficient as in subsubsection 6.3.2 and repeated this process 4 times, each times with a different combination of 3 materials.

Table 4 reports the results of these experiments. Material combinations were randomly chosen, following the procedure described above. The results show Adj. R^2 ranging from 0.75 to 0.81 and MAE ranging from 5.36% and 6.57%, which suggest that it is possible to estimate how different materials would affect gesture recognition performance by considering just 3 reference materials. We can conclude that designers can predict gesture recognition performance for all the materials in the catalog for their own classifiers, provided that the gestures style is similar to the one we have investigated. In the next section we elaborate more on this discussion.

In our software repository we do offer a tool that provides designers with a convenient way to built such linear regression models and visualise the list of material recommendations based on predicted performance, as highlighted in Figure 1(c–d). The tool also includes a simple step-by-step guide, enabling expansion of the catalogue of materials with new materials and thicknesses.

Table 4 Single-predictor linear regression models using Transmission coefficient limited of Energy total (F3) and observations from 3 materials to predict gesture recognition performance on all other materials.

MODEL FIT										MODEL EVALUATION				
Material 1		Material 2		Material 3										
id	name (thickness mm)	id	name (thickness mm)	id	name (thickness mm)	Num materials	DoF	RMSE	R^2	Adj.R^2	Num materials	RMSE	MAE	R^2	Adj.R^2
14	Eva Fome (40)	51	Deb Cerovy.(10)	82	Picea Abies (17)	3	1	0.96	1.00	1.00	72	7.26	5.82	0.81	0.81
42	Silk (<1)	60	Wood Populus (10)	32	Chipboard (32)	3	1	0.25	1.00	1.00	72	8.43	6.11	0.75	0.75
10	MDF (10)	21	Ceramic tiles (6)	75	Dywall (24)	3	1	9.83	0.91	0.82	72	7.59	5.36	0.80	0.80
35	Paper (10 sheets)	79	Paper (200 sheets)	80	Paper (300 sheets)	3	1	1.79	1.00	0.99	72	8.04	6.57	0.79	0.78

Gesture classifier:
Gesture set: [G0 G1 G2 G3 G4 G5]
Dataset: Robot
Material: none
Recordings: 1,200

8 DISCUSSION

Our work can inform application designers interested in estimating a-priori performance of materials for mid-air gesture interaction. The results for *Oriented strand board*, *Paper*, and *Drywall* (section 5) provide evidence for the validity of our proposed approach. Our visual observations and subsequent analysis of core features as time series data shows a high correlation between material thickness and both peak-to-peak amplitude and SNR (Figure 7). Furthermore, our analysis with high-correlation predictors (mean -0.92 for Amplitude and mean -0.82 for SNR) confirms our visual observations. We can conclude that our proposed indirect measurement method is suitable for describing signal distortions as the radar signal passes though various materials. However, our results also indicate that the proposed method fails to accurately describe signal distortions for materials that are transparent to the radar signal, such as *Styrofoam*, *Polyethylene*, *Acrylic*, and *Polycarbonate*. One confounding factor can be attributed to the fact that we tested only thicknesses up to 5 cm for these materials, for which small changes in amplitude and SNR are difficult to detect by our experimental setup. This could be addressed by testing more thicknesses, however, the operational range of mm-wave radar is bound to close proximity (up to 30 cm).

To make our evaluation of the catalogue agnostic to the data source (robot or human) and gesture recogniser, we tested 14 different classifiers: The results show a good fit in all cases, which allows us to conclude that our catalogue can be used for predicting gesture recognition performance for untested materials (unseen data) and similar gesture sets. However, it is important to note that we only evaluated our catalogue on one deep learning model. As it is a state-of-the-art and highly accurate model, we argue that there is no need to test more complex models, unless the gestures are dramatically different from the ones we used.

As the ultimate test, we have shown how to conveniently estimate gesture recognition performance when only characteristics of signal distortions for a few materials are known. We have shown that this is possible to achieve by fitting a simple linear regression model based on performance estimates of just 3 materials from our catalogue (e.g. Adj. R^2 ranging from 0.75 to 0.81 and MAE ranging from 5.36% and 6.57%). We argue that further improvements are possible if we consider more reference materials, but at the same time we acknowledge that it would take additional effort to do so. Overall, our method provides a reasonable ballpark in a matter of minutes.

Our catalogue currently holds 75 materials of various types and thicknesses. There are many more materials and thicknesses one could envision to be used in radar sensing applications. One such example are filaments for 3D printing (e.g. ABS, PLA, PVA, ASA). Due to the low complexity of our experimental setup, one could easily expand the existing catalogue. In principle, this is limited to those users who have access to the Soli sensor and the SDK. However, it is important to note that this does not limit the usage of our catalogue. According to Soli's processing pipeline [Lien et al. 2016], the extracted core features are hardware-agnostic. As such, it should be possible to predict gesture recognition performance for untested materials using our catalogue also with other radar-on-chip sensors which operate in the same frequency range as Soli. We believe this would make for an interesting future work.

Another avenue for future work would be to use a vector network analyser and extend our catalogue with direct measurements of material properties. Researchers with access to such equipment could use the catalogue to further analyse the influence of signal distortion on gesture recognition performance. Perhaps most interesting would be to investigate distortions in signal phase, which is not possible to measure with our proposed method. Signal phase is important when modeling distortion in electronics [Agilent 1997] hence it may hold useful information to predict gesture recognition performance as well.

9 CONCLUSION

We have studied how different materials placed between a mm-wave radar sensor and a human or robot hand would affect gesture recognition performance. Our proposed methodology of using indirect measurements of radar signal properties, requiring only a radar-on-chip sensor and optionally a DIY robot system for automating gesture articulation, is suitable for understanding signal degradation as it passes through various materials. As a result, we have compiled a catalogue of everyday materials that can support gesture designers to determine which materials and gesture sets will perform best in their particular situation. Finally, our results show that it is possible to predict gesture recognition performance on any material similar to the ones we have analyzed. Our catalogue is diverse enough to cover a reasonably large range of materials, so we hope others will find it useful and build upon our work.

REFERENCES

Fadel Adib, Chen-Yu Hsu, Hongzi Mao, Dina Katabi, and Frédo Durand. 2015. Capturing the Human Figure through a Wall. *ACM Trans. Graphics* 34, 6 (2015).

Fadel Adib, Zach Kabelac, Dina Katabi, and Robert C Miller. 2014. 3D tracking via body radio reflections. In *Proc. NSDI*. 317–329.

Agilent 1997. Understanding the Fundamental Principles of Vector Network Analysis. Agilent AN 1287-1.

Shaikh Shawon Arefin Shimon, Courtney Lutton, Zichun Xu, Sarah Morrison-Smith, Christina Boucher, and Jaime Ruiz. 2016. Exploring non-touchscreen gestures for smartwatches. In *Proc. CHI*. 3822–3833.

Francisco Bernardo, Nicholas Arner, and Paul Batchelor. 2017. O soli mio: exploring millimeter wave radar for musical interaction. In *Proc. NIME*. 283–286.

Amit Bleiweiss, Dagan Eshar, Gershom Kutliroff, Alon Lerner, Yinon Oshrat, and Yaron Yanai. 2010. Enhanced interactive gaming by blending full-body tracking and gesture animation. In *ACM SIGGRAPH ASIA 2010 Sketches*. 1–2.

Ke-Yu Chen, Kent Lyons, Sean White, and Shwetak Patel. 2013. uTrack: 3D input using two magnetic sensors. In *Proc. UIST*. 237–244.

Victor C Chen, Fayin Li, S-S Ho, and Harry Wechsler. 2006. Micro-Doppler effect in radar: phenomenon, model, and simulation study. *IEEE Trans. Aerosp. Electron. Syst* 42, 1 (2006), 2–21.

Kevin Chetty, Graeme E Smith, and Karl Woodbridge. 2011. Through-the-wall sensing of personnel using passive bistatic wifi radar at standoff distances. *IEEE Trans. Geosci. Remote Sens.* 50, 4 (2011), 1218–1226.

Gabe Cohn, Daniel Morris, Shwetak Patel, and Desney Tan. 2012. Humantenna: using the body as an antenna for real-time whole-body interaction. In *Proc. CHI*. 1901–1910.

Barrett Ens, Aaron Quigley, Hui-Shyong Yeo, Pourang Irani, Thammathip Piumsomboon, and Mark Billinghurst. 2017. Exploring mixed-scale gesture interaction. In *ACM SIGGRAPH Asia 2017 Posters*. 1–2.

Luigi Gallo, Alessio Pierluigi Placitelli, and Mario Ciampi. 2011. Controller-free exploration of medical image data: Experiencing the Kinect. In *Proc. CBMS*. 1–6.

Brook Galna, Gillian Barry, Dan Jackson, Dadirayi Mhiripiri, Patrick Olivier, and Lynn Rochester. 2014. Accuracy of the Microsoft Kinect sensor for measuring movement in people with Parkinson's disease. *Gait & posture* 39, 4 (2014), 1062–1068.

Nils Y. Hammerla, Shane Halloran, and Thomas Plotz. 2016. Deep, Convolutional, and Recurrent Models for Human Activity Recognition Using Wearables. In *Proc. IJCAI*. 1533–1540.

Sungjae Hwang, Myungwook Ahn, and Kwang-yun Wohn. 2013. MagGetz: customizable passive tangible controllers on and around conventional mobile devices. In *Proc. UIST*. 411–416.

Bryce Kellogg, Vamsi Talla, and Shyamnath Gollakota. 2014. Bringing gesture recognition to all devices. In *Proc. NSDI*. 303–316.

Jungsoo Kim, Jiasheng He, Kent Lyons, and Thad Starner. 2007. The gesture watch: A wireless contact-free gesture based wrist interface. In *Proc. ISWC*. 15–22.

Sven Kratz and Michael Rohs. 2009. HoverFlow: expanding the design space of around-device interaction. In *Proc. MobileHCI*. 1–8.

Christine Kühnel, Tilo Westermann, Fabian Hemmert, Sven Kratz, Alexander Müller, and Sebastian Möller. 2011. I'm home: Defining and evaluating a gesture set for smart-home control. *Int. J. Hum. Comput. Stud.* 69, 11 (2011), 693–704.

David R Large, Kyle Harrington, Gary Burnett, and Orestis Georgiou. 2019. Feel the noise: Mid-air ultrasound haptics as a novel human-vehicle interaction paradigm. *Appl. Ergon.* 81 (2019), 102909.

Mathieu Le Goc, Stuart Taylor, Shahram Izadi, and Cem Keskin. 2014. A low-cost transparent electric field sensor for 3d interaction on mobile devices. In *Proc. CHI*. 3167–3170.

Luis A. Leiva, Matjaž Kljun, Christian Sandor, and Klen Čopič Pucihar. 2020. The Wearable Radar: Sensing Gestures Through Fabrics. In *Proc. MobileHCI*. 1–4.

Yi Li. 2012. Hand gesture recognition using Kinect. In *Proc. CSAE*. 196–199.

Jaime Lien, Nicholas Gillian, M Emre Karagozler, Patrick Amihood, Carsten Schwesig, Erik Olson, Hakim Raja, and Ivan Poupyrev. 2016. Soli: Ubiquitous gesture sensing with millimeter wave radar. *ACM Trans. Graphics* 35, 4 (2016), 1–19.

Gan Lu, Lik-Kwan Shark, Geoff Hall, and Ulrike Zeshan. 2012. Immersive manipulation of virtual objects through glove-based hand gesture interaction. *Virtual Reality* 16, 3 (2012), 243–252.

Jess McIntosh, Mike Fraser, Paul Worgan, and Asier Marzo. 2017a. DeskWave: Desktop Interactions Using Low-Cost Microwave Doppler Arrays. In *Proc. CHI EA*. 1885–1892.

Jess McIntosh, Asier Marzo, Mike Fraser, and Carol Phillips. 2017b. Echoflex: Hand gesture recognition using ultrasound imaging. In *Proc. CHI*. 1923–1934.

Andre Mewes, Bennet Hensen, Frank Wacker, and Christian Hansen. 2017. Touchless interaction with software in interventional radiology and surgery: a systematic literature review. *Int. J. Comput. Assist. Radiol. Surg.* 12, 2 (2017), 291–305.

Pranav Mistry and Pattie Maes. 2009. SixthSense: a wearable gestural interface. In *ACM SIGGRAPH ASIA 2009 Art Gallery & Emerging Technologies: Adaptation*. 85–85.

Pavlo Molchanov, Shalini Gupta, Kihwan Kim, and Kari Pulli. 2015. Short-range FMCW monopulse radar for hand-gesture sensing. In *Proc. RadarCon*. 1491–1496.

Adiyan Mujibiya, Xiang Cao, Desney S Tan, Dan Morris, Shwetak N Patel, and Jun Rekimoto. 2013. The sound of touch: on-body touch and gesture sensing based on transdermal ultrasound propagation. In *Proc. ITS*. 189–198.

Ismail Nasr, Reinhard Jungmaier, Ashutosh Baheti, Dennis Noppeney, Jagjit S Bal, Maciej Wojnowski, Emre Karagozler, Hakim Raja, Jaime Lien, Ivan Poupyrev, et al. 2016. A highly integrated 60 GHz 6-channel transceiver with antenna in package for smart sensing and short-range communications. *IEEE J. Solid-State Circuits* 51, 9 (2016), 2066–2076.

J.Y.H. Ng, M. Hausknecht, S. Vijayanarasimhan, O. Vinyals, R. Monga, and G. Toderici. 2015. Beyond short snippets: Deep networks for video classification. In *Proc. CVPR*.

Kenton O'Hara, Gerardo Gonzalez, Abigail Sellen, Graeme Penney, Andreas Varnavas, Helena Mentis, Antonio Criminisi, Robert Corish, Mark Rouncefield, Neville Dastur, et al. 2014. Touchless interaction in surgery. *Commun. ACM* 57, 1 (2014), 70–77.

Francisco J. Ordóñez and Daniel Roggen. 2016. Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition. *Sensors* 16, 1 (2016).

Michael Otero. 2005. Application of a continuous wave radar for human gait recognition. In *Proc. SPIE*, Vol. 5809. 538–548.

Joseph Paradiso, Craig Abler, Kai-yuh Hsiao, and Matthew Reynolds. 1997. The magic carpet: physical sensing for immersive environments. In *Proc. CHI EA*. 277–278.

Joseph A Paradiso. 1999. The brain opera technology: New instruments and gestural sensors for musical interaction and performance. *J. New Music Res.* 28, 2 (1999), 130–149.

François Petitjean, Alain Ketterlin, and Pierre Gançarski. 2011. A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recognit.* 44, 3 (2011), 678–693.

Carl A Pickering, Keith J Burnham, and Michael J Richardson. 2007. A research study of hand gesture recognition technologies and applications for human vehicle interaction. In *Proc. IET Automotive Electronics*. 1–15.

M PourMousavi, M Wojnowski, R Agethen, R Weigel, and A Hagelauer. 2013. Antenna array in eWLB for 61 GHz FMCW radar. In *Proc. APMC*. 310–312.

Qifan Pu, Sidhant Gupta, Shyamnath Gollakota, and Shwetak Patel. 2013. Whole-home gesture recognition using wireless signals. In *Proc. MobiCom*. 27–38.

Tauhidur Rahman, Alexander T Adams, Ruth Vinisha Ravichandran, Mi Zhang, Shwetak N Patel, Julie A Kientz, and Tanzeem Choudhury. 2015. Dopplesleep: A contactless unobtrusive sleep sensing system using short-range doppler radar. In *Proc. Ubicomp*. 39–50.

Siddharth S Rautaray and Anupam Agrawal. 2011. Interaction with virtual game through hand gesture recognition. In *Proc. IMPACT*. 244–247.

Richard Hartless Richard Rudd , Ken Craig , Martin Ganley. 2014. *Building Materials and Propagation*. Technical Report September. 40 pages.

Andreas Riener, Alois Ferscha, Florian Bachmair, Patrick Hagmüller, Alexander Lemme, Dominik Muttenthaler, David Pühringer, Harald Rogner, Adrian Tappe, and Florian Weger. 2013. Standardization of the in-car gesture interaction space. In *Proc. AutomotiveUI*. 14–21.

Marco Roccetti, Gustavo Marfia, and Angelo Semeraro. 2012. Playing into the wild: A gesture-based interface for gaming in public spaces. *J. Vis. Commun. Image Represent.* 23, 3 (2012), 426–440.

Christian Sandor and Hiraku Nakamura. 2018. SoliScratch: A Radar Interface for Scratch DJs. In *Proc. ISMAR-Adjunct*. 427–427.

R. W. Schafer. 2011. What Is a Savitzky-Golay Filter? [Lecture Notes]. *IEEE Signal Process. Mag.* 28, 4 (2011), 111–117.

Mario Schmidt. 2008. The Sankey Diagram in Energy and Material Flow Management - Part II: Methodology and Current Applications. *Journal of Industrial Ecology* 12, 2 (2008), 173–185.

Gözel Shakeri, John H Williamson, and Stephen Brewster. 2018. May the force be with you: Ultrasound haptic feedback for mid-air gesture interaction in cars. In *Proc. AutomotiveUI*. 1–10.

Merrill Ivan Skolnik et al. 1980. *Introduction to radar systems*. Vol. 3. McGraw-hill New York.

Jie Song, Gábor Sörös, Fabrizio Pece, Sean Ryan Fanello, Shahram Izadi, Cem Keskin, and Otmar Hilliges. 2014. In-air gestures around unmodified mobile devices. In *Proc. UIST*. 319–329.

Peng Song, Wooi Boon Goh, William Hutama, Chi-Wing Fu, and Xiaopei Liu. 2012. A handle bar metaphor for virtual object manipulation with mid-air interaction. In *Proc. CHI*. 1297–1306.

Thad Starner, Jake Auxier, Daniel Ashbrook, and Maribeth Gandy. 2000. The gesture pendant: A self-illuminating, wearable, infrared computer vision system for home automation control and medical monitoring. In *Proc. ISWC*. 87–94.

Ivan E Sutherland. 1964. Sketchpad a man-machine graphical communication system. *Simulation* 2, 5 (1964), R–3.

Ivan E Sutherland. 1965. The ultimate display. In *Proc. IFIP Congress*. 506–508.

Wouter Van Vlaenderen, Jens Brulmans, Jo Vermeulen, and Johannes Schöning. 2015. Watchme: A novel input method combining a smartwatch and bimanual interaction. In *Proc. CHI EA*. 2091–2095.

Klen Čopič Pucihar, Christian Sandor, Matjaž Kljun, Wolfgang Huerst, Alexander Plopski, Takafumi Taketomi, Hirokazu Kato, and Luis A. Leiva. 2019. The Missing Interface: Micro-Gestures on Augmented Objects. In *Proc. CHI EA*.

Qian Wan, Yiran Li, Changzhi Li, and Ranadip Pal. 2014. Gesture recognition for smart home applications using portable radar sensors. In *Proc. EMBS*. 6414–6417.

Saiwen Wang, Jie Song, Jaime Lien, Ivan Poupyrev, and Otmar Hilliges. 2016. Interacting with soli: Exploring fine-grained dynamic gesture recognition in the radio-frequency spectrum. In *Proc. UIST*. 851–860.

Yazhou Wang and Aly E Fathy. 2011. Micro-Doppler signatures for intelligent human gait recognition using a UWB impulse radar. In *Proc. APSURSI*. 2103–2106.

Teng Wei and Xinyu Zhang. 2015. mTrack: High-precision passive tracking using millimeter wave radios. In *Proc. MobiCom*. 117–129.

Huiyue Wu and Jianmin Wang. 2012. User-defined body gestures for TV-based applications. In *Proc. ICDH*. 415–420.

LI Yang, Jin Huang, TIAN Feng, WANG Hong-An, and DAI Guo-Zhong. 2019. Gesture interaction in virtual reality. *Virtual Reality & Intelligent Hardware* 1, 1 (2019), 84–112.

Hui-Shyong Yeo, Gergely Flamich, Patrick Schrempf, David Harris-Birtill, and Aaron Quigley. 2016. RadarCat: Radar Categorization for Input & Interaction. In *Proc. UIST*. 833–841.

Renyuan Zhang and Siyang Cao. 2018. Real-time human motion behavior detection via CNN using mmWave radar. *IEEE Sens. Lett.* 3, 2 (2018), 1–4.

Xu Zhang, Xiang Chen, Wen-hui Wang, Ji-hai Yang, Vuokko Lantz, and Kong-qiao Wang. 2009. Hand gesture recognition and virtual game control based on 3D accelerometer and EMG sensors. In *Proc. IUI*. 401–406.

Chen Zhao, Ke-Yu Chen, Md Tanvir Islam Aumi, Shwetak Patel, and Matthew S Reynolds. 2014. SideSwipe: detecting in-air gestures around mobile devices using actual GSM signal. In *Proc. UIST*. 527–534.

M. Zhao, T. Li, M. A. Alsheikh, Y. Tian, H. Zhao, A. Torralba, and D. Katabi. 2018. Through-Wall Human Pose Estimation Using Radio Signals. In *Proc. CVPR*. 7356–7365.

Yan Zhuang, Chen Song, Aosen Wang, Feng Lin, Yiran Li, Changzhan Gu, Changzhi Li, and Wenyao Xu. 2015. SleepSense: Non-invasive sleep event recognition using an electromagnetic probe. In *Proc. BSN*. 1–6.