THEME ARTICLE: COMPUTATIONAL MATERIALS

Sensing Social Behavior With Smart Trousers

Sophie Skach and Patrick G. T. Healey, *Queen Mary University of London, London, E1 4NS, U.K.*
Rebecca Stewart, *Imperial College London, London, SW7 2BX, U.K.*

Nonverbal signals play an important role in social interaction. Body orientation, posture, hand, and leg movements all contribute to successful communication, though research has typically focused on cues transmitted from the torso alone. Here, we explore lower body movements and address two issues. First, the empirical question of what social signals they provide. Second, the technical question of how these movements could be sensed unintrusively and in situations where traditional methods prove challenging. To approach these issues, we propose a soft, wearable sensing system for clothing. Bespoke "smart" trousers with embedded textile pressure sensors are designed and deployed in seated, multiparty conversations. Using simple machine learning techniques and evaluating individual and community models, our results show that it is possible to distinguish basic conversational states. With the trousers picking up speaking, listening, and laughing, they present an appropriate modality to ubiquitously sense human behavior.

NONVERBAL COMMUNICATION VIA CLOTHES

Textiles are a material we have been familiar with for thousands of years, often in the form of clothing. Seen as an extension of our skin, we use it to culturally and socially express ourselves. The fabrics worn on the body can therefore be understood as tools of nonverbal communication. Turned into a sensing surface, they capture a large range of bodily cues that are part of such communication. Together with other cues like gestures, gaze, and posture, nonverbal behavior makes up a significant part of human interaction and contains detailed information about the nature of a conversation.

From spatial arrangements and body orientation alone, it is possible to identify people's engagement and interpersonal relationships.[1,2] We use gestures, head movement, and postural shifts to manage speaker turns, mark topic shifts, signal attitude, affect or health related behavior.[2–5] Most of these signals derive from the torso. There is little work on legs as interactionally relevant body parts, despite indications that they are rich in social cues too. Postures such as leg crossing or stretching can be signals of perceived behaviors and emotions.[2]

Here, we draw attention to the lower body and explore its role in interaction, expanding work on social signals. To do so, we have designed a wearable sensing system that can capture nonverbal behavior unintrusively: bespoke trousers with embedded textile pressure sensors. Our basic research question is, what information about a conversation can we infer from textiles on the lower body? Can they help to investigate previously overlooked nonverbal cues?

Capturing Social Signals

Sensing sometimes subtle nonverbal cues can be challenging. Many techniques require the physical space to be instrumented. The most common approach to capturing nonverbal behavior in human interaction are camera-based systems. However, the reliance on visual cues is vulnerable to problems with occlusion and can provoke privacy concerns.

On-body sensing avoids this and can be more selective about what data are collected. Ubiquitous approaches have been used to sense affective states with pressure sensors,[4] group dynamics with accelerometers,[6] or support interactional engagement with radio frequency tags.[7,8] Many such sensors and recording systems employ conspicuous forms of

Digital Object Identifier 10.1109/MPRV.2021.3088153
Date of publication 12 July 2021; date of current version 30 July 2021.

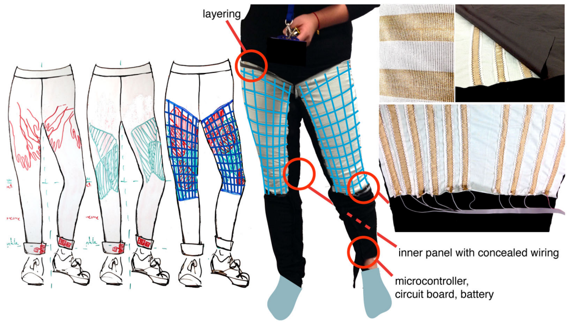

FIGURE 1. *Left:* Findings of ethnographic observations—most common touch surface areas of hands (red) on thighs and of leg crossing postures (green), and the resulting sensor distribution around thighs and buttocks (blue grid). *Middle:* Final prototype of our "smart trousers," turned inside out to show the sensor matrix layer and electronic components. *Right:* The different materials used in the sensor—gray piezo-resistive stretch fabric; conductive stripes of "Zebra" Fabric; and black nonconductive viscose jersey for the trouser shell. *Bottom Right:* Embroidered wiring along inside panels of inner legs.

industrial design, e.g., encapsulated in plastic or integrated in other rigid gadget-like devices, such as wristbands, watches, or belts.[9] This material augmentation such wearable computing systems entail can be overcome with the use of a less intrusive modality: textiles. They provide beneficial properties to sense bodily data and advantages compared to other sensing systems: textiles are soft, flexible, and comfortable to wear on the body. From book covers to car seats to our underpants, they are omnipresent in our environment.

Textile Sensing Systems

As a body-centric sensing surface, textiles have been established as a fundamental part of wearable computing for many decades, used for a variety of applications including healthcare,[9,10] sports,[11] or performing arts,[12,13] identifying gestures, torso movement,[14] sitting postures,[15] and even micromovements like shoulder lifts or breathing.[16] Piezoresistive pressure sensors have proven particularly useful when capturing body posture.[12,15,17]

"Posture-aware" clothing mostly tracks upper body postures,[14] with few explorations toward placing textile sensors on the lower body.[11,18] This parallels the literature on studying social behavior. Moreover, smart textiles are broadly used for egocentric approaches, and are less investigated as a methodology for capturing interaction between humans.[7] With clothes "woven" into embodied social interaction, it is surprising they have not been exploited to a larger extent to capture nonverbal behavior.

The system presented here addresses these gaps and captures basic conversational states using sensors made entirely from textiles with bespoke tailoring techniques.

DESIGNING SENSING TROUSERS

We designed bespoke "smart" trousers, fabricated with custom made textile pressure sensors to measure postural movement during social encounters. Ethnography-based textile sensor design is married with pattern cutting methods from tailoring to design an ideal wearable sensing system, presenting a novel approach of textile sensor integration.

Ethnographic Observations

To develop the sensors and the garment embedding them, we conducted a series of ethnographic observations of seated multiparty conversations, identifying patterns of postural movement that correlate with different behaviors.

Special attention was given to lower body postures, examining which areas are most commonly touched, traces of hand and leg touch illustrated in Figure 1, and observed speaker and listener movement informed the choice of placement, and shape of sensors. With more hand touch on the upper thighs, a denser sensor distribution is required there compared to the side of the legs. The overview of behavioral movement findings is shown in Table 1, indicating the appropriate type of sensor and which behaviors displayed distinct postural movement.

Textile Sensor Design

The observations determine that detecting pressure is a reasonable means to sense postural behaviors on the lower body. The sensing system thus needs to detect both the amount of pressure and the location where it occurs—a task suited to a 2-D matrix of pressure sensors.[12,17]

Matrix Design

Two textile pressure matrices, one on each leg, are constructed using three conductive layers. The top and bottom layers are conductive strips of material with each layer arranged perpendicular to the other. Between the two layers is a single sheet of piezoresistive fabric, which decreases its electrical resistance when pressure is applied.

The sensor designed for the trousers uses 1 cm stripes to form a 10×10 matrix, creating 100 data

TABLE 1. Overview of the coding scheme for the annotated behavioral cues, as well as a summary of ethnographic observations of embodied social behaviors and postural patterns of speakers and listeners.

Class	Subclasses	Description for Annotations	Observed Body Postures
Speaker		Verbal utterance; onset of speaking; overt speech	sitting up straight; no overt movements; distinct gesturing; fewer postural adjustments overall
Active Listener	Backchannels	Verbal response to speaker; repair initiation, e.g., "uhm," "yeh," "ah," etc.	leg crossing; rubbing thighs; more frequent postural changes; more shoulder movement
	Laughter	Verbal and "embodied" laughter; no 'social' smile	occasional lifting of thighs
	Nodding	Distinct up- and downwards movement of head; no separate head turns	as response to speaker when being addressed
Incidental Listener		"Silence"; no distinct listener behavior; includes shoulder shrugging, coughs; other head movement; posture changes; scratching; other 'inattentive,' unspecified listener movement	leaning back; more side activities and overt postural changes

points around each leg (200 in total). The matrix is positioned around the thighs and the buttocks and the sensors' distribution mapped according to the density of common touch points around these areas.

Materials

The trousers consist of three types of material. For the outer shell and legs, a black cotton single jersey knit fabric is used, which ensures a high wearing comfort, elasticity, and good washability. The lining of the trousers, which forms the pressure sensor matrix, consists of three combined layers. The two outer layers of the pressure matrix are constructed from conductive stripes of single jersey knit, cut and sewn onto nonconductive jersey, see Figure 1. The conductive rows and columns of the matrix are 1 cm wide strips of knitted silver-plated nylon thread. A layer of piezoresistive fabric is placed between the rows and columns layers.

Up until this point, only textile materials are used in the trousers, but rigid materials are introduced in order to record the change in pressure from each crossing point of the matrix. The rows and columns of the textile matrix are connected to a Teensy 3.2 microcontroller with thin and flexible insulated wires that are embroidered onto the fabric on one end, and soldered to connectors attached to the circuit board (PCB) on the other. The circuit design and microcontroller code is adapted from Donneaud *et al.*[12] All rigid electronics are concealed in the hem of the trousers.

Tailoring Smart Trousers

Trousers that are to be tested with a large number of different people need to fit each person and fulfill standards of wearing comfort. Both are requirements for the pattern construction of the trousers and are supported by the choice of fabrics that are used. Additionally, the "seamless" and unintrusive integration of the sensors influences the engineering of a basic trouser pattern block.

Pattern Development

A standard pattern block for trousers made of stretch fabric can be seen in Figure 3. The side seam was removed to enable an easier integration of the sensor matrix around the thighs, for which a continuous flat surface was preferable. Housing the microcontroller and 20 wires per leg linking the matrix' rows and columns with the circuit board unintrusively adds certain requirements to the construction of the trousers. On the inside of the legs, a tubular panel was integrated, so all wiring could be pulled through it down to the hem, where it was attached to the circuit board. This panel, manufactured in a technique similar to a "french seam" in tailoring jargon, prevents the electronic components from coming into direct contact with the skin as it is concealed between layers of the soft nonconductive fabric.

The stretch fabric used accommodates multiple clothing sizes with one pattern cut. To allow for a wide range of sizes, however, a grading system was developed, and three sizes of trousers were manufactured, following a process common in the tailoring industry.

Validation of the Embedded Sensors

Fabricating your own sensor out of a flexible, deformable material requires careful testing before deploying a wearable sensing system in an uncontrolled environment. The reliability and overall performance of our

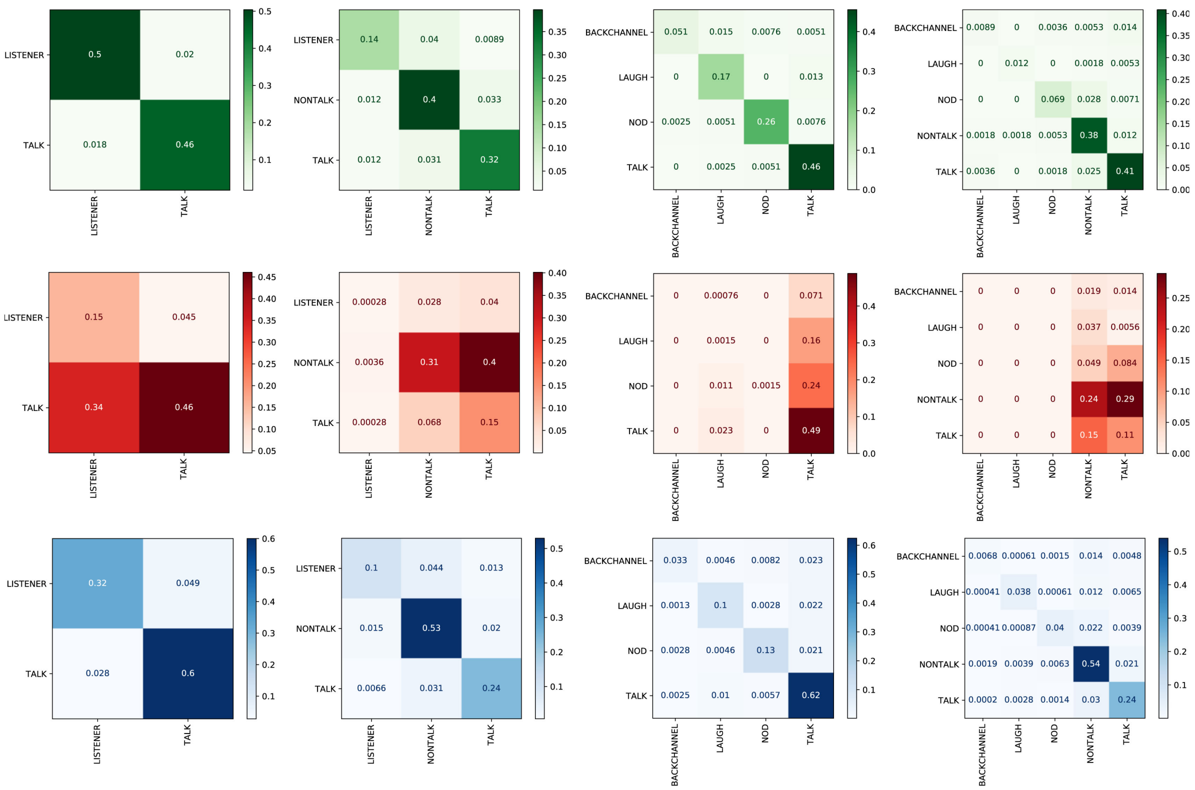

FIGURE 2. Confusion matrices of the Random Forest classification for individual, withheld and community model (from top to bottom) for 2, 3, 4, and 5 class discrimination. The scale for all matrices is the proportion of all instances. (Note the confusion matrices show the results of the classifier weighting each class equally, not balancing the weights to account for smaller dataset classes.).

trousers has been validated in pilot studies with single users wearing the trousers and performing instructed movements. Hereby, we determined the appropriate further data processing and analysis.

EVALUATION

Assuming that different conversational states correlate with postural signatures, we must assure that the sensing system we use is able to distinguish between a variety of such postures. Here, we capture the richness of micromovements and seemingly "invisible" shifts of pressure that may be important indicators for social behaviors. To investigate this further, we have conducted an ethics committee approved user study to evaluate whether the changes in pressure detected by the trousers can be correlated with social behaviors exhibited during a seated conversation.

Participants

We recruited a total of 42 participants to record 14 three-way conversations that took place in all possible gendered arrangements. The data evaluated here stems from a subset of 20 participants, 13 female, and 7 male between 20 and 45 years. All sizes of the trousers are represented.

Procedure

The groups of three were sat around a circular table to encourage equal rights to participate.[1] Each group was given the same conversational task—a moral dilemma—to discuss and resolve between themselves. All conversations lasted between 15 and 25 min.

Data Collection

Throughout the duration of the conversation, the raw data from the 200 sensor points of each pair of trousers was recorded with 10 bits of resolution at 4 Hz. This results in 800 measurements (or 400 per leg) per second, which we refer to as one reading or one instance.

In addition to collecting the pressure sensor data from each of the three pairs of trousers, the 14

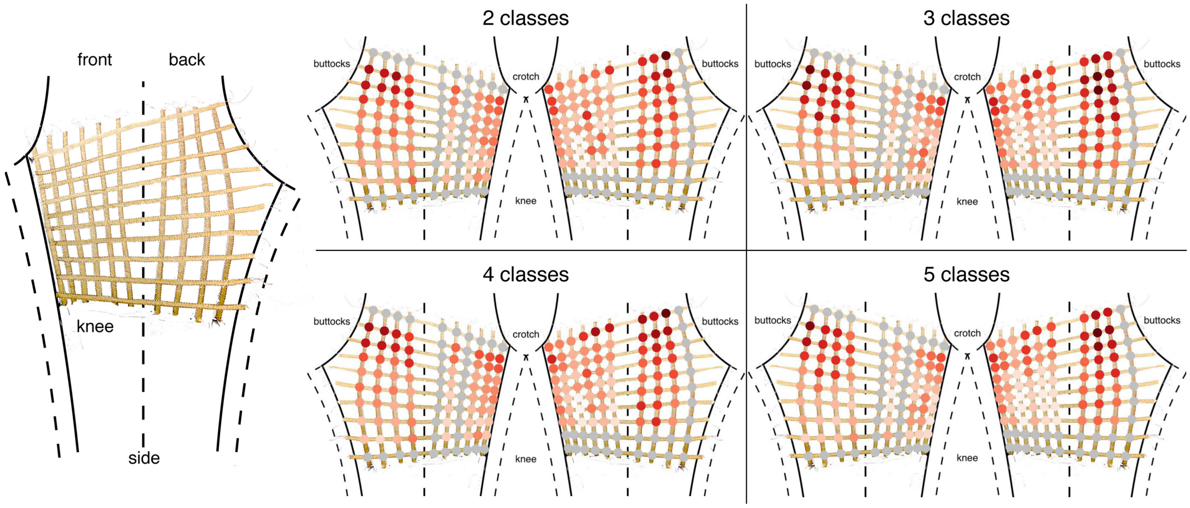

FIGURE 3. Sensor importance of *community* models from all 20 participants for all 2–5 class discrimination variations. The grayed out circles illustrate the sensors not used for analysis (a total of 64 malfunctioning sensors, 28 on the right, 36 on the left leg). The visualization shows the pattern construction of the trousers and distribution of the sensors around the front and back legs. The dashed line shows where the inner panel conceals the wiring.

sessions were recorded with two video cameras that were placed in different corners of the room to capture each participant from various angles.

Coding Behavioral Cues

The video recordings were annotated for four different behavioral states: talking, backchanneling, laughing, and nodding. These were previously selected as basic conversational states and behaviors displaying marked bodily movement, following our observational findings. For modes of talking, we focused on utterances determined through the on and offset of speech. The other modes we distinguish are attentive listener behaviors. Backchannels are identified as verbal responses and initiated repair. When coding for laughter, only concurrent laughter was accounted for, and for nodding is determined by distinct up- and downward head movement. All other movements count toward unspecified behaviors, and were included as a fifth mode that we determine as incidental movement. The definitions of these behaviors can be seen in Table 1.

Each video was hand coded by at least two annotators, and a set of annotation rules was established to identify an annotation with correct starting and end points. Cooccurrences of two behaviors were removed for analysis purposes, and annotations furthermore synchronized with the sensor data of both legs.

Sensor Data Preprocessing

The raw sensor data from the pressure sensor matrix was normalized before any further processing. The time stamp of each leg's sensor data was recreated to be merged with the annotations' timeline. In the scope of this data collection, we encountered two issues: broken sensors and imbalanced datasets.

Any sensor data from malfunctioning sensors was removed before further analysis. In order to evaluate the same set of sensors for each participant, any sensors that broke for one participant's data, were removed from all others'. The final distribution of remaining sensors across both legs of the trousers is visualized in Figure 3.

Based on the hand coded annotations for all identified behaviors, the datasets for each of them were of different size. This derives from the fact that these behaviors occur more or less frequently in relation to each other. Durations of talk are naturally longer than the more brief listener responses (backchannels) or nods. Processing such imbalanced datasets, independent of the analysis techniques to be used, can be handled in different ways. The results reported below stem from an analysis maintaining all of the collected data without modifying the size of datasets and without removing or synthetically adding data of any class. Issues that derive from this approach are accounted for in the classifying methods and address potential problems of imbalanced datasets.

RESULTS

The analysis first explores whether the trousers' pressure sensors can discriminate between the two most basic conversational states: speaking and listening. It then explores whether it can discriminate between three basic states of speaking, active listening, and "incidental movement" of participants not overtly displaying recipiency. Second, we explore whether the trousers' pressure sensors could discriminate between the nonverbal response directed movements characteristic of active listeners (backchannels, nods, and laughter), first without including incidental movements and then with incidental movements. This resulted in testing for the ability to automatically discriminate between the following behaviors:

- 2 classes: Speaker and Active Listener cues;
- 3 classes: Speaker, Active Listener, and Incidental Listener cues;
- 4 classes: Speaker, Backchannels, Laughter, and Nodding cues;
- 5 classes: Speaker, Backchannels, Laughter, Nodding, and Incidental Listener cues.

Furthermore, the ability to discriminate between these classes was examined at the individual level for each of the 20 participants and at the community level for a generalized model representative of all participants.

In total, the dataset consists of 22 870 instances of talking gathered from all 20 participants and 12 095 instances of active listener behaviors (backchannels + nods + laughter). Among active listener behaviors, backchannels had the fewest instances (2380—equivalent to ca. 10 min), followed by laughter (4383) and nods (5062). The distribution of cues derives from the duration and frequency of occurrence of each of the behaviors. Measures were taken to compensate for the imbalanced number of instances.

Classifier Model Selection

There is a variety of classification algorithms that can be used and have previously been used in connection with smart textiles and social signal processing.

Four types of models to distinguish between the behaviors were initially investigated: Support vector machines (SVM), K-nearest neighbour (KNN) algorithm, Gaussian Naive Bayes (GNB), and random forests. Each of them bears different advantages and disadvantages in regards to the type of data we work with, and by testing the different models, we explored these characteristics before selecting a random forest classification to carry further and analyze our data with. We will elaborate on random forest in detail below, and give a brief summary of the results of the other classifiers here. For all tests, the data were kept in its original, imbalances format and split into a training (60%) and test set (40%).

The SVM and GNB showed the poorest results. While overall average accuracies for individual and community models appear good especially for 2 and 3 class discriminations, examining the F-Measures demonstrates the weakness of the two models for our data and classification task. The KNN achieved promising overall accuracies for all classes.

In comparison, a random forest classifier outperformed the above for both, individual and community models for all different social behavior analyses. Based on these performances, we focus here on reporting only the results of random forest. The particular model we evaluate here uses a fivefold cross validation with stratified data and bagging with 100 iterations. The trees are built with unlimited depth. Considering the imbalanced datasets we are evaluating against each other here, the different classes are weighted inversely proportional to how frequently they appear in the overall dataset.

Speakers and Listeners

Two sets of models were trained and evaluated using fivefold cross validation: one to discriminate between the two classes of speakers and active listeners only, and a second set of models to discriminate between three classes, adding incidental listening. These are the instances where neither active listening nor speaking behaviors are exhibited.

First, each participant was treated as an independent dataset and an individual model was trained and evaluated using fivefold cross validation. Then, the aggregate dataset of all participants was used to train a model also evaluated using fivefold cross validation. Last, 20 models trained with 19 participants were evaluated against the withheld participant.

Individual Models

The best mean accuracy across all individual models is the one discriminating between two classes, both with equally weight distribution and balanced weigh assignment based on the size of datasets. Also, Precision, Recall, and F1 Scores (F-Measures) are high, averaging between 0.91 and 0.93. The F1 Measures are 0.86 for listeners and 0.96 for speakers, averaged across all participants, as shown in Table 2.

TABLE 2. F1 Measures of the random forest (RF) classification per class, averaged across individuals. below are the RF classification results for individual participants, community models, and withheld participants across all classes.

	2 classes	3 classes	4 classes	5 classes
F1 Measures for Individuals per class				
Talk	0.958	0.856	0.976	0.865
Incidental Listener	–	0.940	–	0.946
Active Listener	0.865	0.628	–	–
Backchannels	–	–	0.519	0.263
Nodding	–	–	0.812	0.549
Laughter	–	–	0.788	0.667
Individual Participants				
Accuracy	0.932	0.879	0.904	0.872
Balanced Accuracy	0.912	0.810	0.770	0.660
Precision	0.933	0.866	0.868	0.808
Recall	0.912	0.810	0.770	0.660
F1 Measure	0.919	0.830	0.798	0.701
Community Model				
Accuracy	0.931	0.877	0.898	0.867
Balanced Accuracy	0.920	0.825	0.783	0.677
Precision	0.931	0.868	0.878	0.832
Recall	0.920	0.825	0.783	0.677
F1 Measure	0.925	0.843	0.821	0.729
Withheld Participants				
Accuracy	0.521	0.463	0.615	0.457
Balanced Accuracy	0.534	0.337	0.251	0.197
Precision	0.534	0.337	0.237	0.186
Recall	0.534	0.332	0.251	0.196
F1 Measure	0.488	0.284	0.200	0.164

All results are average outputs of the classifier across participants.

When the third class for incidental listeners is included, the overall mean accuracy drops. This also reflected in Recall, Precision, and F-Measure results across all 20 participants.

Moreover, including incidental listening leads to more variation between listeners and speakers, too, see Table 2. An overview of the average results of Recall, F1 Measures, Precision, and the two calculated accuracies are listed in Table 2.

Looking at misclassifications, for both 2 and 3 class discriminations, speakers, and listeners are rarely mixed up. Expectedly, incidental listener movement is rather misclassified as listeners than it is as speakers. Figure 2 also reveals that adding this additional class does not necessarily decrease the performance of the model.

Community Models

Applying the same cross validation with 100 trees to a general community model, the results are of similarly high accuracies, with 93.1% correctly classified instances for the 2 class discrimination, and a lower 87.7%, and similarly when balancing the weight so that smaller datasets are assigned more weight, see Table 2.

Examining the confusion matrices of these community models, we can see that speaking always performs slightly better than active listening, which yet shows good average Precision, Recall, and F-Measures. Talk is proportionally rarely misclassified, even in the 3 class scenario. Both community model confusion matrices are illustrated in Figure 2 with plotted normalized results.

Leaving One Out

Each participant was tested against a community level when being withheld from the training set, which consisted of the data of the remaining 19 participants. Here, the average percentage of correct classification present a notable decrease in overall performance compared to a general community model, and is sometimes just above chance. Modifying the weight of the differently sized datasets, the mean balanced accuracy is slightly better for the 2 class discrimination, but slightly worse when including incidental listening. The average results for Precision, Recall, and F1 Measures can be seen in Table 2.

We can also compare the withheld participant's performance with the individual model, as is presented through the normalized confusion matrices in Figure 2. This shows that even the participant with the best results in the individual model does not keep up when tested against the community model, but shows results around the average. Other participants that performed among the best in the individual models did not have better results when being withheld from the training set.

Arguably, the number of instances of the test set could be responsible for the variety of results and the overall weak performance of withheld participants. In cases of larger test sets, the performance overall was better. On the other hand, a participant with a lower number of training instances and a therefore higher number of test instances shows more extreme Recall results for each class: very low active listener and relatively high speaker results. But these examples are only marginal appearances and do not represent the overall behavior of the community model.

Backchannels, Laughter, Nods

Next we explore whether the textile pressure sensors can not only distinguish listeners from speakers, but also more fine-grained conversational states. The same training and evaluation procedure used to evaluate the discrimination between speakers and active listeners is now applied to 1) discriminating between the subclasses of active listeners and then to 2) discriminating between speakers, the subclasses of active listeners, and incidental listeners through the addition of sensor data of the unspecified "silence," determined by the gaps of all other coded behaviors.

Individual Models

For the 4 class individual model, the overall average percentage of correct classifications is 90.4%, and 77.0% when balancing the weight distribution, while for the 5 class model, it is 87.2%, but only 66.0% for a balanced accuracy, compare Table 2. Like in the previous groupings of behaviors, this drop in results was expected given that incidental listening entails all unspecified movement and nonverbal signals.

Both, for the 4 and 5 behavior discrimination, Precision, Recall, and F-Measure results demonstrate that among the differentiated active listener behaviors, laughter performs best, followed by nodding, which shows slightly better F1-Measures for 4 classes than for 5.

In comparison, however, talk outperforms the active listener behaviors by far, also including the most diverse signals and movements. In the 5 class scenario, incidental listening scores highest.

Examining confusion matrices of the 4 class model, as well as Recall and F-Measure, we see that while "talk" performs best, nodding and laughing also show good results and rarely misclassify each other. Most participants with high Recall and F-Measure results for laughter and nods, also have above average results for backchannels. This outcome could be an indicator for the ability of the system to detect fine grained differences of behavior, while it struggles more to compare those against a more generic state. These results are illustrated with the confusion matrix of one representative participant in Figure 2 (top).

The confusion matrices reveal similar insights in the 5 class discrimination, yielding talk as the strongest and best performing category. This additional behavioral category shows a wider spread of misclassifications across the remaining four, but has overall least mix ups with speaking.

Community Models

The general community model for the discrimination between the specified 4 behaviors shows an overall performance with Recall results of 0.777 for backchannels, 0.833 for laughter, 0.809 for nods, and 0.706 for talk, see Table 2, this reflects the good results of the individual models, too.

The confusion matrix of the 4 class model in Figure 2 (bottom) shows that all active listener behaviors are only on rare occasions confused with each other, but rather with talk, which itself seems to be more distinct to nods than to laughter with fewer misclassifications toward this class.

Including incidental listener movement yields the lowest results for Precision, Recall, and F1 Measures, so that the previously strong class of talking performs weaker.

Leaving One Out

When training the community model for 19 participants and test it on the withheld one, the results are of poor precision in both occasions. We can further see a large difference in accuracies when applying the different weighing of datasets to the classifier, as Table 2 reveals. For the withheld participants, the misclassification results for the different behaviors vary a lot with variations of the data size. In the 4 class discrimination, the best Recall results are 0.875 for backchannels, 0.544 for laughter and 0.124 for nods, while talk performs much better with up to 0.988. Backchannels and nods present mostly the smallest sample sets among all behavioral cues, both resulting in the poorest F1 measures.

The worst results are drawn from the 5 class discrimination for withheld participants. This can also be observed in the confusion matrix in Figure 2 (middle), which displays the results of a single, representative participant. Almost all misclassifications happen toward the biggest datasets—talking and incidental listening.

Feature Importance

In addition to the predictions the Random Forest classifier provides, we evaluate the importance of each sensor across the pressure matrices in both of the trousers' legs. The feature vector is built from the sensor input and is based on the gini impurity. This feature extraction helps to gain a better understanding of the classifier, and to reduce the high dimensionality of the 200 sensors.

The visualization of sensor importance shows that the area around the upper buttocks is most significant for all classes. The different legs of the trousers appear to be of slightly asymmetric importance, and the front outer thighs of lowest importance overall, as Figure 3 shows. On this first glance, there is a major overlap of sensor importance across all four classification models. When examining the results in detail, however, we can find fine grained differences between the four scenarios. In particular, there seems to be a shift in significant sensors when including the class of incidental listening. For example, the upper buttocks area appears more relevant for the 2 and 4 classes only distinguishing between speakers and listeners, while the mid or central buttocks are more relevant when incidental listening is included. As for the sensors covering the thighs, finally, the sensors in the crotch area seem to only distinguish between the 2 class scenario and the rest. Differences between speakers and active listeners yield the inner top thigh close to the crotch area less important than in other multiclass models.

Extracting the important features for each participant reveals a relatively large individual variation across all multiclass discriminations. Some sensors and sensor groups of seemingly low importance are specific to one participant appear important for others. Compared to the community model's feature importance, we can observe a minor trend toward slightly higher sensor importance on the inner leg, as well as on the outer, toward the side seam leaning buttocks area, rather than on mid or inner buttocks, and the top thigh in general.

DISCUSSION

Overall, the results show that we can use wearable textile pressure sensing systems in clothing to detect basic conversational states. Our findings introduce trousers as a suitable ubiquitous sensing surface that helps us explore social signals the lower body transmits—something previously left largely unattended.

Additional Cues and Classification Features

The questions around detecting nonverbal cues the lower body provides focused on three active listener signals, but there are more to exploit. The variety of "incidental" movements can be split into further subclasses like fidgeting or axis related postures, potentially improving classification accuracies. Another parameter affecting the performance of the models is time. Including this in the analysis, and collecting more data, other methods like neural networks can be explored.

Moreover, detecting changes in pressure distribution can also be used to identify different addressees and interpersonal correlations, as well as topic changes. It has been shown that information on postural movement can even be used to detect chronic pain behavior.[5] With trousers as a method to capture nonverbal behavior, however, some signals remain undetected, such as facial and gestural micromovements.

Even Smarter Trousers

The visualization of the feature importance yields groups of sensors that are more significant than others to discriminate the determined classes, sparking discussions as to how many or few sensors are needed to detect behavioral cues. Future iterations can be designed to address this aspect, reducing the amount of sensors significantly and optimizing their placement. For example, examining the sensor importance draws attention to the buttocks as a relevant area for detecting social signals. As one of largest muscles and link between the upper and lower body, they can be used to explore a wide range of nonverbal cues expanding on the ones discussed previously.

Additionally, design engineering parameters like the resolution of the recording frequency, the robustness of hard-soft connections, and additional manufacturing techniques such as[19] allowing for an ever better merging of layers and processes can be reviewed and optimized for ever smarter trousers.

Individual Variation

Both, in the confusion matrices as well as the feature importance, a large individual variation is observed. While general community models that average the dataset of all 20 participants used in this analysis, show good performances, the F1 Measures drop when withholding one participant from the training set and testing on the individual. This indicates that when aiming to detect signals of unknown individuals, our system is not ready to be deployed in its current state. The good performance of the individual models across all classes, however, lets us conclude that people show repetitive nonverbal patterns, but everyone does so differently. This characteristic can support trousers that are to identify their wearers, for example, where their own sample data is required, not that of others.

CONCLUSION

Our approach of "socially aware clothing" explores postural movement that proves challenging to detect with more traditional technologies, and contributes to a discussion of use cases for smart textiles in a social context. We have introduced a sensing system to capture these signals without modifying our surroundings—smart trousers with embedded textile pressure sensors. This is a step toward establishing textiles as a novel, wearable sensing system for applications in social sciences, and contribute toward a better understanding of nonverbal communication.

In this work, the different groupings of behaviors show excellent results for individual participants and on a general community level. It is only when withholding a participant in the training set, the boundaries of our sensing system are shown. While there are further measures to address this, one conclusion we can draw from these results is, that each of us has developed distinct signals of embodied social behavior that we perform consistently, but that can be very different from other conversation partners. We are, in this sense, more similar to ourselves than to others. And while this leads to poor results in the testing model we report here, it also bears advantages in regards to personalized, social computing applications.

ACKNOWLEDGMENTS

We thank all annotators and participants for their time; our colleagues and friends for their support, and the reviewers of this paper for their valuable suggestions. This work was supported by EPSRC and AHRC Centre for Doctoral Training in Media and Arts Technology under Grant EP/L01632X/1.

REFERENCES

1. A. Kendon, *Conducting Interaction: Patterns of Behavior in Focused Encounters*. Cambridge, U.K.: Cambridge Univ. Press, 1990.
2. P. Bull and G. Connelly, "Body movement and emphasis in speech," *J. Nonverbal Behav.*, vol. 9, no. 3, pp. 169–187, 1985.
3. P. G. Healey and S. A. Battersby, "The interactional geometry of a three-way conversation," in *Proc. 31st Annu. Conf. Cogn. Sci. Soc.*, 2009, pp. 785–790.
4. S. S. D'Mello, P. Chipman, and A. Graesser, "Posture as a predictor of learner's affective engagement," in *Proc. Annu. Meeting Cogn. Sci. Soc.*, vol. 29, no. 29, 2007, pp. 905–910.
5. C. Wang, T. A. Olugbade, A. Mathur, A. C. D. C. Williams, N. D. Lane, and N. Bianchi-Berthouze, "Recurrent network based automatic detection of chronic pain protective behavior using mocap and semg data," in *Proc. 23rd Int. Symp. Wearable Comput.*, Sep. 2019, pp. 225–230.
6. H. Hung, G. Englebienne, and J. Kools, "Classifying social actions with a single accelerometer," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, Sep. 2013, pp. 207–210.
7. N. Khaorapapong and M. Purver, "Designing unobtrusive interfaces to increase naturalness of first time face-to-face interaction," in *Proc. Int. SERIES Inf. Syst. Manage. Creative eMedia*, 2013, pp. 19–28.
8. T. Choudhury and A. Pentland, "Sensing and modeling human networks using the sociometer," in *Proc. 7th IEEE Int. Symp. Wearable Comput.*, Oct. 2003, pp. 216–222.
9. S. Patel, H. Park, P. Bonato, L. Chan and M. Rodgers, "A. review of wearable sensors and systems with application in rehabilitation," *J. Neuroeng. Rehabil.*, vol. 9, no. 1, pp. 1–17, 2012.
10. K. Arquilla, A. K. Webb, and A. P. Anderson, "Textile electrocardiogram (ECG) electrodes for wearable health monitoring," *Sensors*, vol. 20, no. 4, 2020, Art no. 1013.
11. R. R. Manero *et al.*, "Wearable embroidered muscle activity sensing device for the human upper leg," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2016, pp. 6062–6065.
12. M. Donneaud, C. Honnet, and P. Strohmeier, "Designing a multi-touch etextile for music performances," in *Proc. 17th Int. Conf. New Interfaces Musical Expression*, May 2017, pp. 7–12.
13. M. Satomi and H. Perner-Wilson, "Kobakant DIY Wearable Technology Documentation," 2007. [Online]. Available: https://www.kobakant.at/DIY/
14. A.García Patiño, M. Khoshnam, and C. Menon, "Wearable device to monitor back movements using an inductive textile sensor," *Sensors*, vol. 20, no. 3, 2020, Art. no. 905.
15. J. Meyer, B. Arnrich, J. Schumm, and G. Troster, "Design and modeling of a textile pressure sensor for sitting posture classification," *IEEE Sensors J.*, vol. 10, no. 8, pp. 1391–1398, Aug. 2010.
16. L. Dunne *et al.*, "Garment-based body sensing using foam sensors," in *Proc. 7th Australian User Interface Conf.*, Hobart, Tasmania, Jan. 16–19, Jan. 2006.
17. P. Strohmeier, V. Håkansson, C. Honnet, D. Ashbrook, and K. Hornbæk, "Optimizing pressure matrices: Interdigitation and interpolation methods for continuous position input," in *Proc. 13th Int. Conf. Tangible, Embedded, Embodied Interact.*, Mar. 2019, pp. 117–126.

18. L. E. Dunne, G. Gioberto, V. Ramesh, and H. Koo, "Measuring movement of denim trousers for garment-integrated sensing applications," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2011, pp. 3990–3993.
19. K. Klamka, R. Dachselt, and J. Steimle, "Rapid iron-on user interfaces: Hands-on fabrication of interactive textile prototypes," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2020, pp. 1–14.

SOPHIE SKACH is currently a Postdoctoral Researcher with Queen Mary University of London, London, U.K. She received the Ph.D. degree with the School of Electronic Engineering and Computer Science as a member of the Media & Arts Technology Programme. Trained as a fashion designer, her research interests combine studies of nonverbal social behavior and the exploration of textile sensing systems on the body in interactional context. Contact her at s.skach@qmul.ac.uk.

REBECCA STEWART is currently a Lecturer with the Dyson School of Design Engineering, Imperial College London, where she leads the e-Body Lab. Her research encompasses wearable technology with a focus on audio and textile interfaces and also examines how to communicate technical research findings to applications designers. Contact her at r.stewart@imperial.ac.uk.

PATRICK HEALEY is a Professor of Human Interaction and the Head of the Cognitive Science Research Group , School of Electronic Engineering and Computer Science, Queen Mary University of London, London, U.K. He is a fellow of the Turing Institute and of the Society for Text and Discourse. He recently served as a Senior Researcher in Residence with the Digital Catapult and as International Visiting Chair in Empirical Foundations of Linguistics at Sorbonne Cité/Paris 7. His research interests focus on miscommunication in human interaction. Contact him at p.healey@qmul.ac.uk.