Human Activity Recognition using a mmWave Radar

GIES DEN BROEDER, University of Twente, The Netherlands

Human Activity Recognition (HAR) is becoming increasingly useful for applications such as well-being monitoring and personalizing smart spaces. Traditional methods for HAR often require wearable devices or camera's. The former is not feasible for every environment and the latter has strong privacy concerns. The mmWave radar has been shown to be a promising alternative as it does not imply the same privacy concerns and does not require the users to have wearable devices. In this paper we have used a low-cost mmWave radar to generate micro-Doppler spectrograms to ultimately classify different activities. For this, multiple classifiers and methods of spectrogram filtering have been examined. Finally a Time-Distributed Convolutional Neural Network in conjunction with a Bi-Directional Long Short-Term Memory has attained an average accuracy of 99.62% on a dataset of 5 activities, involving 2 participants.

Additional Key Words and Phrases: Activity Recognition, mmWave Radar, micro-Doppler, Spectrogram, machine learning

1 INTRODUCTION

Knowing the location of people and what they are doing can be of importance to a variety of applications. Elderly require constant monitoring to allow them to live an independent lifestyle, while also ensuring their well-being [1]. Smart spaces can better respond to personalized demands, such as heating, lighting, security management and sound selection using this information, increasing comfort and energy efficiency [8]. Currently user, identification and activity tracking methods include using visual camera's, WiFi and device-based solutions, where users are identified by their smartphone, watch or ID-card [11]. While camera's achieve great performance in these tasks, they do have the downsides of light-condition reliance, as well as privacy concerns. Camera's are intrusive and often poorly received in both domestic and commercial settings [2]. In addition, camera's in a hospital have been used to spy on female patients [5]. WiFi-based solutions require a separate transmitter and receiver, and only work when the target is located between them, limiting their usability. Moreover, device-based solutions require human effort and assume inseparability of the device and their users, properties that are ultimately undesired for seamless integration.

The mmWave radar is a small device, operating as a transceiver using electromagnetic waves. In addition its waves can penetrate thin layers of some materials, allowing it to be placed inside furniture or walls [6]. These properties can make the mmWave radar a better fit for user, identification and activity tracking purposes than the aforementioned solutions.

Thus, in this paper the efficacy of the mmWave radar will be tested. While it can be used for user-tracking, user identification, activity recognition and more, the focus of this paper will be solely on activity recognition. However, the methods used can be applied to the variety of applications discussed above.

2 RELATED WORK

Human activity recognition (HAR) has been researched extensively over the past decade or so. As discussed in Section 1, solutions to HAR include visual camera's, Inertial Measurement Units (IMU) and even WiFi Routers. However, because of the constraints and concerns associated with these types of sensors, the radar-based solutions are most relevant to this research.

Recently, Frequency Modulated Continuous Wave (FMCW) radars have been increasingly used, because the higher frequency allows for superior range resolution. However, instead of outputting raw data, these devices automatically generate a point cloud. The number of points in each frame is not consistent, making it more complex to use the point cloud as input for machine learning classifiers, as the data is not of constant dimensions.

Singh. et al. circumvent this problem by transforming the generated point clouds into a voxelized representation, which makes the data dimensions constant, but also makes each individual frame a large size of 10 x 32 x 32. They manage to achieve accuracies of 90% using deep learning classifiers, showing that automatic feature extraction can be as performant as manual feature extraction using traditional machine learning methods. Zhao et al. also use the generated point cloud, though they use it in real time by clustering the points together, tracking the clusters and finally identifying them [11].

While the point clouds are the usual outputs of the FMCW radars, it is possible to use the raw data of these radars to detect micro-Doppler effects from moving targets. These micro-Doppler effects arise when non-rigid bodies move, as along with the general movement of the body/torso, small micro-scale movements and rotations also occur. Think of the swinging of the arms and the movement of the feet during walking. These micro-Doppler effects can be visualized using a micro-Doppler spectrogram. A big advantage that the spectrogram has over the point clouds is its data size. Using the voxel presentation presented by Singh et al. each frame has a data size of over 10.000 [9]. Meanwhile, the spectrogram usually has a data size of 100 - 300 for a single frame. Additionally, similarly to the point clouds, the spectrogram has also been demonstrated to achieve high accuracy in (activity) recognition tasks.

Kim et al. have extracted features from the spectrogram manually to train an SVM classifiers to achieve an accuracy of 90% for activity recognition [7]. Zhang et al. used it to recognize basic human activities and achieve accuracies of over 90% [10]. Janakaraj et al. used the spectrogram for the purpose of human identification based on people's gait and achieve an accuracy of 97.45% on a dataset of 20 people [3]. Moreover, the spectrogram proves useful for smaller movements as well, as Jiang et al. have used it to detect hand gestures using an SVM and CNN, attaining a best accuracy of 95% [4].

In this work, we have collected a HAR dataset containing 5 activities using a mmWave radar operating in the 77-81 GHz range. Using the

TScIT 37, July 8, 2022, Enschede, The Netherlands

raw data of the radar, micro-doppler spectrograms have been generated for these activities. The spectrograms have been enhanced using filtering methods to remove background noise and have subsequently been used to train different classifiers. The best performing classifier has been able to achieve an average accuracy of 99.62%.
The rest of the paper is organized as follows. In Section 3 preliminary information about the workings of the radar and the signal processing methods is provided. Section 4 mentions the experimental setup, the post-processing methods and the classifiers are presented. Section 5 discusses the results and finally Section 6 concludes the paper.

3 BACKGROUND

3.1 Radar

A mmWave radar is a radar that works in the mmWave range. They operate at frequencies between 30-300 GHz where the waves have a size of 1 - 10 mm, hence the name. For this research, the TI IWR 1443Boost was used, which works in the 77-81 GHz frequency band. It contains 4 receivers and 3 transmitters, allowing it to receive the reflections of the sent signal. Due to having multiple receivers it is also able to determine the Angle of Arrival (AoA) on the horizontal plane of reflected waves.
Additionally, this is also a Frequency Modulated Continuous Wave (FMCW) radar, meaning that the signal of the radar is sent in 'chirps' and 'frames'. Each chirp is a continuous wave, but linearly increases in frequency for the duration of the chirp. The frequency during a chirp is sampled a set amount of times, which is denoted by 'ADC samples'. A frame consists of multiple consecutive chirps and usually ends with a refractory period where no signal is being sent. The range and velocity of an object can be measured using the ADC samples and the chirps, respectively.

The resolution of these properties can be influenced by the radar configuration. These are important, because most activities to be recognized involve micro-movements of peripheral limbs such as the hands, arms or legs. Higher resolutions improve the fidelity of the measurements of such activities and give the machine learning classifiers higher quality data to learn from.

Range resolution (d_{res}) determines the minimum distance between 2 objects, such that they can still be distinguished as different objects. The formula for the range resolution is as follows:

$$d_{res} = \frac{c}{2B} \tag{1}$$

where c is the speed of light and B is the bandwidth of the sweeping chirp. Meaning the total bandwidth of a single chirp is the only factor determining the final range resolution. The frequency band of the IWR 1443Boost allows it a maximum bandwidth of 4 GHz, which would yield a range resolution of 3.75cm.
Velocity resolution (v_{res}) determines the minimum frequency difference between 2 discrete frequencies, such that their sum can be resolved into their respective parts. The formula for the velocity resolution is as follows:

$$v_{res} = \frac{\lambda}{2T_f} \tag{2}$$

where λ is the wavelength and T_f is the time of a single frame. With the radar, there is limited control over the final wavelength, but its high frequency allows for a strong velocity resolution compared to lower frequency radars. However, the frame time can be controlled.

3.2 Signal Processing

Using the raw data of the radar, it is possible to generate a Micro-Doppler Signature (MDS), also known as a (micro-)doppler spectrogram. The MDS is simply a plot of reflected (doppler) frequency or velocity over time. This means that static objects only show up on the x-axis where the velocity or frequency is 0, meaning that there is no possibility to distinguish between non-moving objects. However, the MDS allows micro-movements of peripheral limbs such as the hands, arms and legs to be distinguishable from the bigger moving parts such as the torso. Depending on the radar configuration varying levels of detail can be seen on the MDS.
One possibility to generate the MDS is as follows. First the data must be arranged into the radar data cube. This is a way of arranging the data such that the data has dimensions as follows: (# of Rx Channels, # of ADC Samples, # of Chirps, # of Frames). The # of Frames is simply the time dimension, and is traditionally not included in the radar data cube. This representation allows for Fast Fourier Transforms (FFT) to be easily applied to the correct dimensions. The following steps must be performed over each frame. First the Range-FFT is applied over the ADC samples of a frame. Next, we sum over the same dimension. Secondly, a hanning window is applied over the chirps, after which the Doppler-FFT will be performed. Finally, we sum over the Rx channels. One such MDS can be seen in Figure 1.

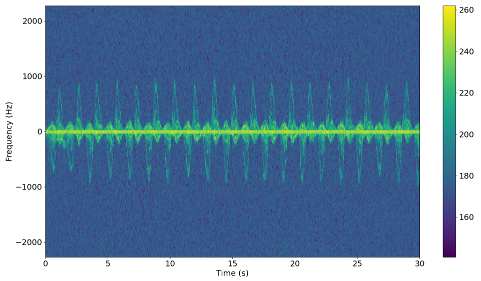

Fig. 1. Raw Micro-Doppler Spectrogram of a Squat

4 METHODOLOGY

4.1 Experimental Setup

In Table 1 the used radar configuration can be found. As outlined in Section 3.1, the range and velocity resolutions are both very important for the results of activity recognition. It was found that range resolution could be improved by increasing the bandwidth of the sweeping chirps, and velocity resolution could be improved by shortening the frame time. The bandwidth increase could have been facilitated by either lengthening the chirp time and/or increasing the chirp slope. However, lengthening the chirp time has consequences for the amount of chirps that can fit in a frame, so a higher chirp slope was chosen. More chirps in a frame increases the resolution of the spectrogram. As such, it was chosen to not decrease the frame

Table 1. Radar Configuration

Parameter	Value
Tx	1
Start Frequency	77 GHz
ADC Samples	256
Chirp Slope	66 GHz/s
Bandwidth	3963 MHz
Chirps per Frame	200
Frames	750
Periodicity	40 ms

Table 2. Activities

Activity	# of Records	Total Duration (s)
Clapping	26	780
Jogging	26	780
Jumping Jacks	26	780
Squats	26	780
Waving	26	780

time from the default settings and instead increase the amount of chirps per frame significantly. Consequently, this does increase the amount of data that needs to be processed. However, compared to increasing the ADC samples to 512 or using 2 transmitters instead of 1, the increase in data size is marginal. Regarding increasing the ADC samples and transmitters used, neither were found to make a significant difference in the spectrograms, and thus the increase in data size was not deemed to be worth it. Finally, 750 frames were chosen to make every measurement 30 seconds long.

For the data collection the radar was mounted on a tripod at a height of 1.2m. The activities were performed at a distance of 2m from the radar. The setup can be viewed in Figure 2. In total, 5 activities were performed by 2 participants. Each participant performed each activity 13 times for 30 seconds. This yields a total of 6.5 minutes of data per activity per person, totalling 65 minutes of data. The activities that were performed can be seen in Table 2, along with the total number of records and total duration of recorded data.

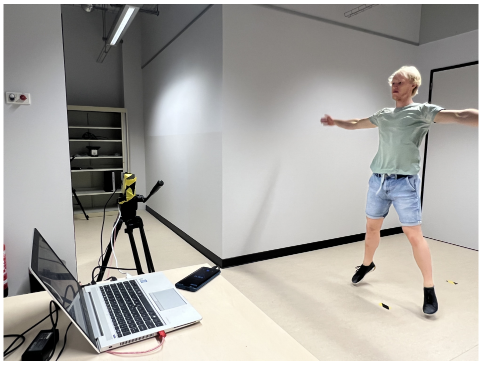

Fig. 2. Data Collection setup

4.2 Post-Processing

Before training the different classifiers, the raw spectrograms need to be improved. Particularly, the background noise needs to be filtered to allow for the best possible results. For this purpose different filtering methods were examined.
Filter 1 simply subtracts the mean of the entire spectrogram from all points in the spectrogram. Subsequently all negative points are set to 0.
Filter 2 uses the steps of filter 1 twice.
Filter 3 first applies a gaussian blur with a sigma value of 0.5 to the spectrogram, after which the same steps from filter 1 are used.
The sigma value of the gaussian blur determines how strong the blur is. A higher sigma value will increase the disparity between background noise and signal, but will blur the signal in the process. In Figure 4 in Appendix A all filters are applied to the same spectrogram as shown in Figure 1. It can be seen that filter 1 retains the strongest signal, though also the most background noise. Filter 2 has a slightly weaker signal, but also has very little background noise left over. Finally filter 3 has very little background noise as well, however the signal is of slightly lower quality due to the gaussian blur. After filtering, each spectrogram is sliced into slices of 2 seconds (50 frames), with an overlap of 0.32 seconds (8 frames). The 2 second time window was chosen based on previous works of human activity recognition and identification [9, 11] This gives 2288 samples per activity, making 11.440 samples in total.

4.3 Classifiers

Different Machine Learning Classifiers have been trained, specifically the Support Vector Machine (SVM), the Long Short-Term Memory (LSTM) and a Convolutional Neural Network (CNN) combined with an LSTM were trained. These classifiers were taken from the Github page [1] of the RadHAR paper [9]. The MLP from their page was also trained, but did not converge past an accuracy of 20% and was decided to not be included in the results. The only difference with their classifiers is the input size, as each spectrogram slice has a size of 50 x 200 whereas each 2 second window of the RadHAR voxel representation has a size of 60 x 10 x 32 x 32. Consequently the CNN classes are now 1D classes instead of 3D classes. Each of these classifiers were trained on the same train-test split of the gathered dataset. A train-test split of 75/25 was used for the training. They were implemented using sklearn and keras. For the deep learning classifiers an Adam Optimizer was used with a learning rate of 0.001. The models were trained for 30 epochs, during which the models with minimum loss were saved.

4.3.1 SVM.

The SVM receives as data input a flattened representation of each slice. Principal Component Analysis (PCA) is used to reduce the dimensionality from 10.000 to 100. GridsearchSVC was used in conjunction with an RBF kernel.

4.3.2 Bi-Directional LSTM.

The Bi-Directional LSTM is a classifier in which the LSTM layer is duplicated, with the first layer receiving the input data from past

[1] https://github.com/nesl/RadHAR

Table 3. Classifiers with Accuracy per Filter

	Accuracy		
Classifier	Filter 1	Filter 2	Filter 3
SVM	95.87%	96.56%	96.77%
LSTM	96.63%	97.65%	97.60%
CNN + LSTM	99.48%	99.62%	99.40%

Table 4. Classifiers with Accuracy Range per Filter

	Range		
Classifier	Filter 1	Filter 2	Filter 3
SVM	95.70 - 96.08	96.33 - 96.92	96.22 - 97.10
LSTM	96.29 - 96.92	97.13 - 98.21	96.64 - 98.32
CNN + LSTM	99.20 - 99.65	99.30 - 99.83	99.09 - 99.55

to future and the second layer receiving it from future to past. This allows the classifier to retain information from both past and future. It consists of the Bi-Directional LSTM layer, followed by 2 fully connected layers, with the output layer as final layer.

4.3.3 *Time-Distributed CNN + Bi-Directional LSTM.*

The Time-Distributed CNN applies a CNN layer to every temporal slice of the input data. The complete model consists of 3 Time Distributed convolutional modules, each including 2 convolution layers and 1 maxpooling layer. Finally it has a Bi-Directional LSTM layer, followed by the output layer. This is the only trained classifier that has the capability to use the spatial dimensions of the input data in addition to the time dimension.

The results of the classifiers can be found in Table 3. The reported accuracies are the average of 4 training session, where each session was done on the exact same train-test split. In Table 4 the ranges of the achieved accuracies are also shown. Figure 3 contains the confusion matrix for the best performing CNN + LSTM model from filter 2. Only waving is not predicted with a 100% accuracy, however it does not confuse enough samples to show what it actually predicted. It would make sense that it predicted clapping instead of waving, as those 2 activities have the smallest total movement and thus the smallest signal in the spectrogram.

5 DISCUSSION

All filtering methods achieve extremely good performance for all classifiers. The high accuracy of the SVM is somewhat surprising, given that in the RadHar paper it only achieved just over 60% on a similar dataset. Without applying a PCA the accuracy was significantly lower. This could be due to removing noise that was leftover from the filtering steps. The Bi-Directional LSTM performs slightly better than the SVM. This classifier has some concept of time and tries to look at the sequence and timing of the input data. Since human activities are usually performed over a short duration, it makes sense that the LSTM performs well. The Time-Distributed CNN + Bi-Directional LSTM performs phenomenally with an accuracy of over 99% for both filters. Given that this is basically the Bi-Directional LSTM with additional convolutional layers before it, it is logical that it performs well. Despite the fact that the input data does not have spatial dimensions, it is still able to learn extremely well.

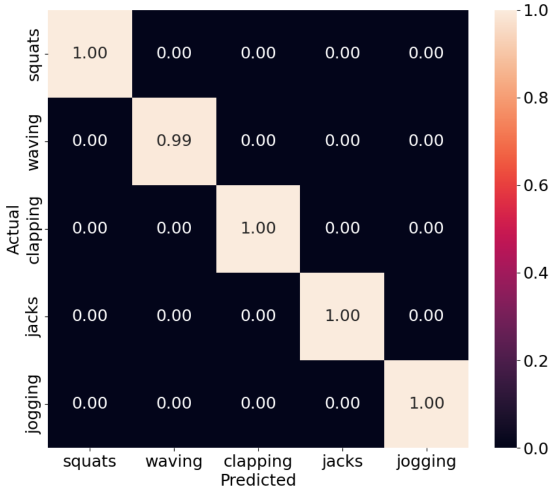

Fig. 3. Confusion Matrix of the best CNN LSTM using Filter 2

There are 2 main reasons that contribute greatly to the high accuracy for this research. Firstly, the chosen activities were relatively different and consequently produce quite different spectrograms as well. Obviously this makes it easier to classify the activities correctly, compared to when more similar activities would have been chosen. Secondly there were only 2 participants that contributed to the dataset. Because people move in different ways, the spectrograms that 2 different people generate by performing the same activity can be very different. Thus if more people were involved in creating the dataset, they would be harder to classify.

The difference in accuracy between the 3 filters is almost negligible. Filters 2 and 3 do perform ever so slightly better than filter 1 for the SVM and LSTM. This could be due to the lower total noise using those filters. The differences for the CNN + LSTM are extremely minor, and so no real conclusions can be drawn from this. However, I would hypothesize that the small increase in accuracy for filter 2 is due to the lower noise level and the slightly lower accuracy for filter 3 has a basis in the reduced quality of the signal.

6 CONCLUSION

In this paper, we generated our own human activity dataset using a low-cost mmWave radar for the purpose of Human Activity Recognition. Using the radar's raw data, micro-Doppler spectrograms have been created and subsequently used to train different machine and deep learning classifiers. Using the classifiers of the RadHAR paper [9] on a similar dataset we have been able to achieve superior results with the best combination of classifier and filtering method achieving an average accuracy of 99.62%. This could imply that micro-Doppler spectrograms are a superior signal processing option compared to the sparse point clouds, both in performance and training data size.

REFERENCES

[1] Ferhat Attal et al. “Physical Human Activity Recognition Using Wearable Sensors”. In: *Sensors 2015, Vol. 15, Pages 31314-31338* 15.12 (Dec. 2015), pp. 31314–31338. issn: 1424-8220. doi: 10.3390/S151229858. url: https://www.mdpi.com/1424-8220/15/12/29858/htm%20https://www.mdpi.com/1424-8220/15/12/29858.

[2] Robert Beringer et al. “The “Acceptance” of Ambient Assisted Living: Developing an Alternate Methodology to This Limited Research Lens”. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 6719 LNCS (2011), pp. 161–167. issn: 03029743. doi: 10.1007/978-3-642-21535-3{_}21. url: https://link.springer.com/chapter/10.1007/978-3-642-21535-3_21.

[3] Prabhu Janakaraj et al. “STAR: Simultaneous Tracking and Recognition through Millimeter Waves and Deep Learning”. In: *2019 12th IFIP Wireless and Mobile Networking Conference (WMNC)*. 2019, pp. 211–218. doi: 10.23919/WMNC.2019.8881354.

[4] Wen Jiang et al. “Recognition of dynamic hand gesture based on mm-wave FMCW radar micro-Doppler signatures”. In: *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings* 2021-June (2021), pp. 4905–4909. issn: 15206149. doi: 10.1109/ICASSP39728.2021.9414837.

[5] John Bonifield. *Cameras secretly recorded women in California hospital delivery rooms - CNN*. 2019. url: https://edition.cnn.com/2019/04/02/health/hidden-cameras-california-hospital/index.html.

[6] David D Ferris Jr. and Nicholas C Currie. “Microwave and millimeter-wave systems for wall penetration”. In: *Targets and Backgrounds: Characterization and Representation IV*. Ed. by Wendell R Watkins and Dieter Clement. Vol. 3375. SPIE, 1998, pp. 269–279. doi: 10.1117/12.327159. url: https://doi.org/10.1117/12.327159.

[7] Youngwook Kim and Hao Ling. “Human activity classification based on micro-doppler signatures using a support vector machine”. In: *IEEE Transactions on Geoscience and Remote Sensing* 47.5 (May 2009), pp. 1328–1337. issn: 01962892. doi: 10.1109/TGRS.2009.2012849.

[8] Chris Xiaoxuan Lu et al. “SCAN: Learning speaker identity from noisy sensor data”. In: *Proceedings - 2017 16th ACM/IEEE International Conference on Information Processing in Sensor Networks, IPSN 2017*. Association for Computing Machinery, Inc, Apr. 2017, pp. 67–78. isbn: 9781450348904. doi: 10.1145/3055031.3055073.

[9] Akash Deep Singh et al. “Radhar: Human activity recognition from point clouds generated through a millimeter-wave radar”. In: *Proceedings of the Annual International Conference on Mobile Computing and Networking, MOBICOM*. Association for Computing Machinery, Oct. 2019, pp. 51–56. isbn: 9781450369329. doi: 10.1145/3349624.3356768.

[10] Renyuan Zhang and Siyang Cao. “Real-Time Human Motion Behavior Detection via CNN Using mmWave Radar”. In: *IEEE Sensors Letters* 3.2 (Feb. 2019). issn: 24751472. doi: 10.1109/LSENS.2018.2889060.

[11] Peijun Zhao et al. “MID: Tracking and identifying people with millimeter wave radar”. In: *Proceedings - 15th Annual International Conference on Distributed Computing in Sensor Systems, DCOSS 2019*. Institute of Electrical and Electronics Engineers Inc., May 2019, pp. 33–40. isbn: 9781728105703. doi: 10.1109/DCOSS.2019.00028.

A FILTERED SPECTROGRAMS

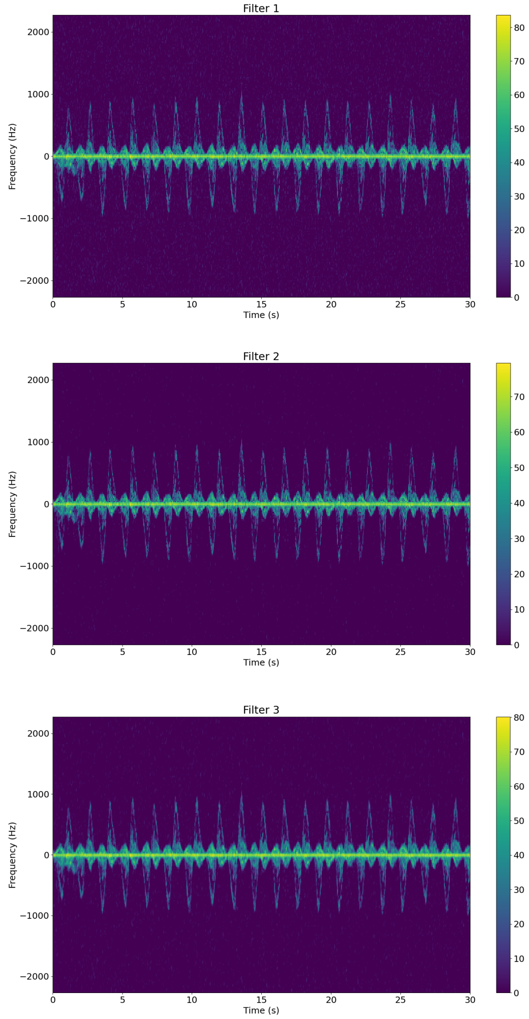

Fig. 4. Filtered Micro-Doppler Spectrograms of a squat