
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Raja, Muneeba; Vali, Zahra; Palipana, Sameera; Michelson, David G.; Sigg, Stephan
3D Head Motion Detection Using Millimeter-Wave Doppler Radar

Published in:
IEEE Access

DOI:
[10.1109/ACCESS.2020.2973957](https://doi.org/10.1109/ACCESS.2020.2973957)

Published: 01/01/2020

Document Version
Publisher's PDF, also known as Version of record

Published under the following license:
CC BY

Please cite the original version:
Raja, M., Vali, Z., Palipana, S., Michelson, D. G., & Sigg, S. (2020). 3D Head Motion Detection Using Millimeter-Wave Doppler Radar. *IEEE Access*, 8, 32321-32331. Article 8998250.
<https://doi.org/10.1109/ACCESS.2020.2973957>

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

Received January 22, 2020, accepted February 5, 2020, date of publication February 13, 2020, date of current version February 24, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2973957

3D Head Motion Detection Using Millimeter-Wave Doppler Radar

MUNEEBA RAJA¹, (Member, IEEE), ZAHRA VALI², SAMEERA PALIPANA¹,
DAVID G. MICHELSON², (Senior Member, IEEE), AND STEPHAN SIGG¹, (Member, IEEE)

¹Department of Communications and Networking, Aalto University, 02150 Espoo, Finland

²Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

Corresponding author: Muneeba Raja (muneeba.raja@aalto.fi)

This work was supported by the CHIST-ERA III, Radiosense Project.

ABSTRACT In advanced driver assistance systems to conditional automation systems, monitoring of driver state is vital for predicting the driver's capacity to supervise or maneuver the vehicle in cases of unexpected road events and to facilitate better in-car services. The paper presents a technique that exploits millimeter-wave Doppler radar for 3D head tracking. Identifying the bistatic and monostatic geometry for antennas to detect rotational vs. translational movements, the authors propose the biscaltering angle for computing a distinctive feature set to isolate dynamic movements via class memberships. Through data reduction and joint time–frequency analysis, movement boundaries are marked for creation of a simplified, uncorrelated, and highly separable feature set. The authors report movement-prediction accuracy of 92%. This non-invasive and simplified head tracking has the potential to enhance monitoring of driver state in autonomous vehicles and aid intelligent car assistants in guaranteeing seamless and safe journeys.

INDEX TERMS Bistatic radar, Doppler effect, head movements, STFT, 3D motion detection, wireless sensing.

I. INTRODUCTION

The role of humans in driving is expected to shift from active control to supervision or to passiveness in advance driver assistance systems (ADAS) to partial and conditional automation transportation. Monitoring driver state is vital for ascertaining the capability of supervising or maneuvering the vehicle when unexpected road events occur and for understanding the driver's comfort level.

In-vehicle sensing of driver attention/distraction can be achieved by analyzing posture changes and head, hand, foot, and eye movements [1], [2]. Sensing of human activities in a vehicular setting has been a hot topic because of its importance in ADAS and conditionally autonomous driving and its potential in facilitating human–car interaction and monitoring driver health/fatigue/distraction in driving-safety systems. Most of the current solutions are based on either visible-light cameras or wearable sensors. Camera-based sensing is affected strongly by variations in light levels during a journey,

and in darkness it relies on external illumination, which could distract the driver. Also, privacy concerns arise [3], [4].

In addition, the technology entails a trade off between frame rate (decreasing motion blur) vs. cost. While wearable sensors [5] can overcome some of these issues, wearing the devices may be inconvenient for the driver, and the connection to the environment (e.g., via the steering wheel) can limit which driver states may be inferred (e.g., stress and fatigue) [6].

In recent years, device-free human sensing from wireless signals [7] has gained popularity, thanks to being less invasive and coping with occlusion and darkness. This development has inspired in-vehicle wireless human sensing in the form of monitoring vital signs, such as heart rate [8] and breathing [9], and the driver's state [10]–[12]. Our WiBot system [10] characterizes driver motion from head turns and hand gestures by means of WiFi signals, and, by developing WiCAR [11], we extended this to humans, in-car models, and external-setting-independent in-vehicle activity recognition. Importantly, all three studies utilized CSI as the signal descriptor, since WiFi is seeing increasing application for in-vehicle entertainment. WiFi's operating frequency, however,

The associate editor coordinating the review of this manuscript and approving it for publication was Haiwen Liu.

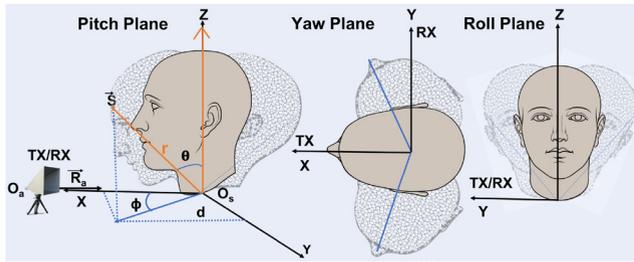


FIGURE 1. Detecting head movements via mmWave signals.

is susceptible to the multipath effect due to passengers and nearby cars, and the head tracking's accuracy is constrained by wavelength. We have addressed this via radio signals at widely separated frequencies (1.8 and 30 GHz) in a monostatic configuration, for distinguishing a driver's head motions from those of the body [13].

So far, most radio frequency sensing efforts have focused on changes in the received signal strength, channel state information, and micro-Doppler motion [14]. In contrast, we have used millimeter-wave Doppler radar to perform in-vehicle driver/passenger 3D head tracking for three distinct head movements, as illustrated in Fig. 1.

Pitch, roll, and yaw movements in 3D space are of great interest because they provide information on driver attention levels and behavior [15]. For example, repeatedly pitching forward and backward may indicate drowsiness, yaw movements may point to shoulder checking and thereby indicate awareness of one's surroundings, and rolling motions may reveal distraction (e.g., reaching for objects or attending to rear-seat passengers).

We recently presented detection of translational movements and separating between the head and torso [13]. In this paper, we demonstrate that appropriate choice of the radar configuration (bistatic vs. monostatic) followed by estimation of the time-varying Doppler spectra allows distinguishing and accurately estimating rotation and translation of the head within a single system, which has not been studied before. While head translation along the radar line of sight leads to a Doppler shift in the frequency of the reflected signals, the rotational head movements have minute impact on the Doppler spectra and introduce a time-varying Doppler spectrum around the carrier frequency. A car-like arrangement in an anechoic chamber with a carrier frequency of 30 GHz was used to investigate the optimal radar configuration of receiver (RX), transmitter (TX), and subject.

A dynamic object's detectability by radar depends on its size, geometry, and wavelength. Size is inversely proportional to wavelength, which means that $\lambda < \text{size of the body part}$ if one is to obtain the complete reradiated signal. At 30 GHz, λ is 1 cm, which is much smaller than a substantial head movement, so these motions can be captured with high accuracy.

Our approach computes a distinctive feature set by marking movement boundaries and performing joint short-term

time–frequency analysis. Head rotations introduce Doppler modulations that can be examined to estimate the manner in which the head is moving. Via reduction and removal of redundancies and correlations in the feature set, we reduce it to the two most separable features. The final step is to separate among movement classes by using a support-vector machine classification algorithm.

The in-vehicle environment is normally static; i.e., the physical surroundings are fixed, and only the human body is in motion. This enables us to focus on dynamic Doppler-based features and to extract features that allow distinguishing between classes of movement, among them a pitch forward, pitch backward, roll left, and roll right (translation) and also a yaw left and yaw right (rotation).

To the best of our knowledge, the 3-dimensional head movement detection using mmWave wireless radar has not been presented before. The only published work closest to our research is [12], that detects 2D head orientation with CSI signals. Besides, this is the first work that present the importance of bistatic and monostatic configuration of TX/RX antennas and the subject, to separate rotational and translational head movements and improve the accuracy of detection. And most importantly, the idea of translating head movements detected from non-intrusive wireless signals to analyze human distracted behavior is a new advancement.

The main objectives of this work are to

- 1) Derive the optimal geometry to separate yaw movements in a car-like arrangement within an anechoic chamber and show that 30 GHz millimeter-wave signals are appropriate for this scenario
- 2) Demonstrate the advantage of bistatic over monostatic configurations for RF sensing that identifies head rotation and translation, since it captures more scattered components of the signal
- 3) Show that translation along the radar line of sight is more easily distinguished in a monostatic configuration and that detecting all rotational- and translational-movement classes within one system requires additional RX units, at 0° and 90° .

The discussion is organized such that Section II describes the concepts that were the building blocks of our research. Details of the experimental setup, human study and frequency–time analysis are presented in Section III, and Section IV covers the system implementation. The results are presented in Section V followed by system limitations in Section VI and conclusions in Section VII.

II. FUNDAMENTAL CONCEPTS

In this section, we introduce the significance of mmWave radar and its specifications in our work, then specify the target head movements and the reasoning behind focusing our methodology on them. Finally, we provide an empirical demonstration of the feasibility of both monostatic and bistatic configurations with regard to translational and rotational movements.

A. THE SIGNIFICANCE OF MMWAVE RADAR

Millimeter-wave radar is a well-established vehicular technology. Short-range radar applications (parking assistance and pre-crash applications), medium-range ones (cross-traffic alerts, lane-change assistance, and blind spots' detection), and long-range ones (adaptive cruise control) already operate in the automotive domain, at such frequencies as 76–81 GHz [16]. Typically, a frequency-modulated continuous waveform is used for deriving information about the existence, location, and velocity of neighboring vehicles from reflections. We are not, however, aware of applications of mmWave radar within the passenger cabin for monitoring driver state, even though it offers significant advantages over lower-frequency regimes with regard to in-vehicle driver state monitoring.

The accuracy with which minute movement can be recognized from RF signals is conditioned on the wavelength for the signal frequency and also on the bandwidth available (for 30–300 GHz, $\lambda=0.001\text{--}0.01$ m). Moreover, in in-car scenarios, the reduced range of mmWave signals does not limit the sensing capabilities, since the distance between antenna and sensed subject is small (usually 1–2 m). Compared to other mmWave spectrum areas, the frequencies around 30 GHz specifically provide a good compromise in this respect, showing the lowest attenuation among all frequencies above 20 GHz [17]. Since device-free RF sensing relies on reflected, heavily attenuated signal components, this makes frequencies around 30 GHz particularly suitable for in-vehicle device-free RF sensing.

B. TARGET MOVEMENT DESCRIPTION

Humans' head movements are physically complex and unstructured but contain rich information about behavioral characteristics. Moving the head relies fundamentally on support against the force of gravity. Head movements steer and are closely linked to the sensory structures of the head, especially those for vision [18]. Humans tend to use a combination of head and eye movements to stabilize their line of sight and focus on a target [19]. Especially in head rotations, the eye movements follow the head to keep gaze shifts in balance. In addition, head movements such as nodding or shaking are used for nonverbal communication. Therefore, head motions, alongside eye movements, are vital in characterizing human states of attention and interest. Since motions of the head have a bigger impact on RF signals than eye movements do, we chose them as our target movement.

In situations wherein the human's position is partially fixed, such as being seated (in a car or otherwise), locomotion is restricted and a subset of complex head movements such as pitch, roll, and yaw could provide enriched information about the behavior, interest, and attention. The emphasis here is on large, significant movements for behavior-detection purposes, and we specify these in the form of class memberships involving the horizontal (pitch and roll) and rotational (yaw) plane. The geometry of these movements is explained below.

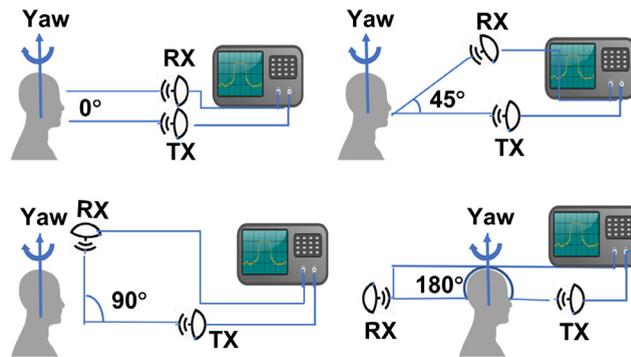


FIGURE 2. The TX–RX configurations used for finding the optimal bistatic angle for accurate yaw detection.

1) TRANSLATIONAL HEAD MOVEMENTS

As depicted in Fig. 1, the subject is modeled as a vector \vec{S} from the origin O_s , pointing toward the head. Relative to O_s , \vec{S} is described in the Cartesian coordinate system as (x, y, z) and in the spherical coordinate system as (r, θ, ϕ) , where $\phi = 0^\circ$ is in the $+x$ -axis direction. The antenna coordinate system is described by (r_a, θ_a, ϕ_a) relative to origin O_a , where $\vec{R}_a = (r_a, 0, 0)$ is the direction of maximum gain. The subject is placed electrically far from the antenna ($\|O_s - O_a\| \gg \lambda$).

For measuring pitch, the receiver is placed such that \vec{R}_a is parallel to the x -axis. Pitch is defined as a change in θ in the $\phi = 0^\circ$ direction,

$$\vec{\omega}_p = (\omega_p \cdot \hat{\theta} + 0 \cdot \hat{\phi} + 0 \cdot \hat{r}) \tag{1}$$

Projecting the angular velocity $\vec{\omega}_p$ onto the x -axis shows that the component of the angular velocity in the direction of \vec{R}_a results in a measurable change in the Doppler frequency at the receiver; therefore, any pitch forward or backward by the head can be detected in changes to the Doppler shift (see Fig. 1). A roll movement with angular velocity $\vec{\omega}_r$ is defined in the same manner as a change in θ in the $\phi = 90^\circ$ direction, and \vec{R}_a is parallel to the y -axis (the receiver is placed beside the subject).

2) ROTATIONAL HEAD MOVEMENTS

Yaw of the head is a rotational movement defined as change in ϕ in the direction of $\theta = 0^\circ$,

$$\vec{\omega}_y = (\omega_y \cdot \hat{\phi} + 0 \cdot \hat{\theta} + 0 \cdot \hat{r}) \tag{2}$$

Yaw movement of the head is expected to cause less Doppler frequency shift than head pitch or roll. This is because yaw movement is partially projected onto the x -axis and does not result in significant components projected onto either the x - or the y -axis. Hence, the change in Doppler frequency is not sufficient for distinguishing between right and left yaw. These can be distinguished, however, with an appropriate bistatic configuration and the joint frequency–time analysis discussed in the following sections.

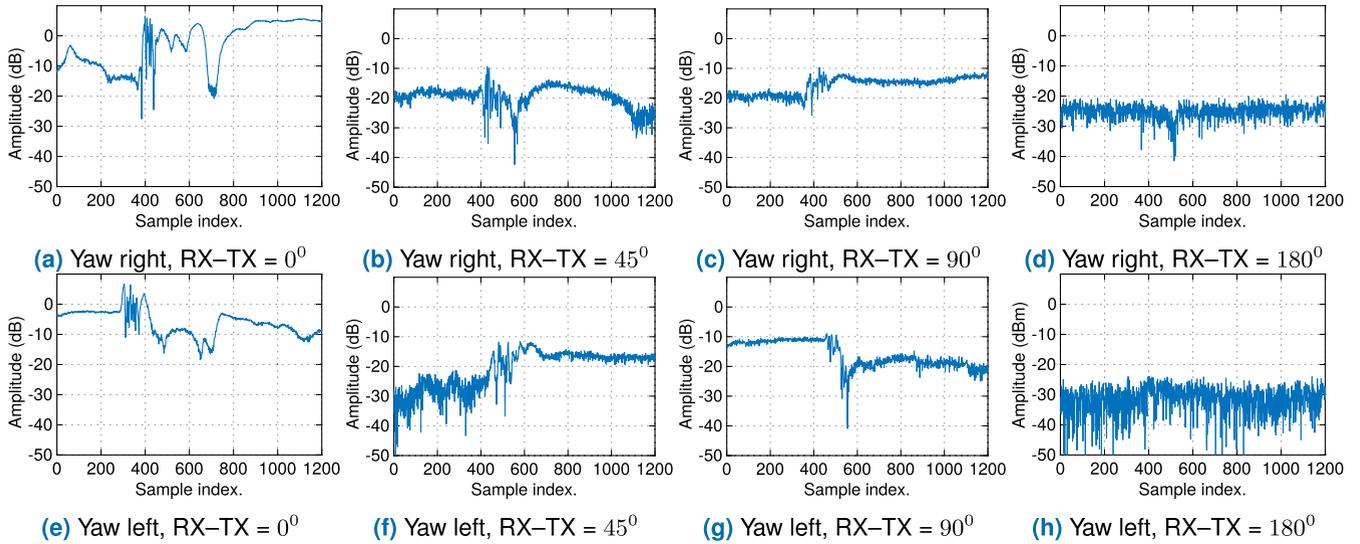


FIGURE 3. Head-rotation-caused fluctuations in normalized RSS values in various system configurations.

C. THE IMPACT OF GEOMETRY ON MOVEMENT DETECTION

While TX and RX are collocated in a monostatic radar configuration, bistatic radar is a system configuration in which they are separated by a considerable distance [20]. The bistatic configuration is specified in terms of the bistatic angle, β , which is defined as the angle (0–180°) between TX and RX, with its vertex at the target. Bistatic radar has complex geometry relative to monostatic radar but provides better receiver sensitivity, a refined radar cross-section, and wider spatial diversity [21].

The different geometry leads to differences between the two setups in scattering of the radiating signal. The angle of incidence between the transmitting and receiving antenna is large in a bistatic setup, in sharp contrast to the collocated antennas in a monostatic one, and this guarantees that the scattering will not be confined to the strongest scattering lobe and should significantly reduce the impact of backscatter [22]. That creates motivation for identifying the less observable or less dominant movements with a reduced radar cross-section (RCS). Scattering due to movement in a different direction results in a large RCS in some bistatic configurations, revealing the subtle movements and providing a higher signal/clutter ratio [23]. This implies that bistatic radar should be favored when the target reflects very little energy in the monostatic direction. Although the design and development of an optimal bistatic geometrical configuration is challenging, it could significantly improve the classification accuracy since this can capture information about those dynamic scattering properties of the target that are invisible to monostatic radar [24], [25]. The bistatic system’s geometry affects radar operation characteristics such as the Doppler equation. For a static transmitter and receiver, the bistatic Doppler frequency is given by f_B [26]

$$f_B = \frac{2V}{\lambda} \cdot \cos(\delta) \cdot \cos(\beta/2), \tag{3}$$

where β is the bistatic angle, δ is the angle between the target velocity vector and bistatic angle bisector, and V is target velocity. We consider different β (0°, 45°, 90°, and 180°) as shown in Fig. 2, and keep all experiment settings and parameters (described in Section III) constant to find out how changing β can affect the signal characteristics from head movements.

D. THE IMPACT OF MOVEMENTS ON THE RSS

The RSS is the variation in amplitude of the base signal caused by human body movements, which is the parameter most easily measured by mainstream wireless technologies.

Our first step was to analyze the impact of head rotations on the RSS. The normalized values for RSS indicator graphs for yaw left and yaw right in various configurations are given in Fig. 3. These graphs provided some fundamental information about the configurations that aided in Doppler effect and STFT analysis. The 180° configuration does not capture any movement information and has high noise levels (see Fig. 3d and 3h), the reason being the subject sitting in the TX and RX line of sight and blocking all the signal that should be received at the RX. The 90° configuration displays a clear gradient rise and drop for yaw right and left (see Fig. 3c and 3g). This occurs because rotational head movement leads to multipath scattering, and less-observable movement is captured more in the bistatic configuration as explained in Subsection II-C. The 90° configuration seems the most reasonable in comparison to 45° and 0°. Head movement is more easily detected with a larger radar cross section (RCS) which depends on the target’s reflectivity, i.e., the RCS of the head is unique in each configuration. In 90° configuration, the complete head rotation from TX to RX gets captured while the complete range of motion for 90° head rotation is not covered at 0° and 45° configurations, refer to Fig. 3a, 3e, and 3b, 3f.

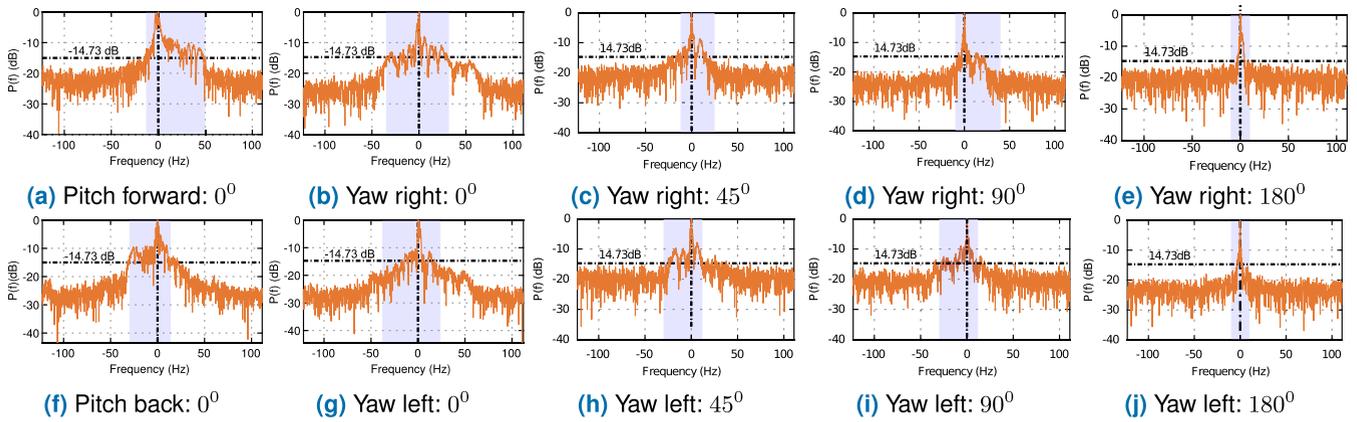


FIGURE 4. Fluctuation in Doppler spread values in various configurations.

E. THE IMPACT OF MOVEMENTS ON THE DOPPLER EFFECT

The Doppler effect is the primary feature used in detecting body movements. A positive frequency shift indicates that the target is moving toward the RX, while a negative one indicates moving away from the RX. The full distribution of the frequency shift due to the Doppler effect is called the Doppler spread, B_D . We calculated B_D by computing the fast Fourier transform (FFT) of the received signal [27].

Fig. 4a and 4f show the Doppler spread graph for pitch movements, toward and away from the RX in the monostatic configuration. Both forward and backward pitch motions are clearly distinguishable from the Doppler spread, since the Doppler shift is significantly greater in one direction for each movement. Yaw right and left movements have a less pronounced impact and Doppler shift at either end of the frequency spectrum and create almost equal shift on the two sides so cannot be separated with a monostatic configuration (cf. Fig. 4b and 4g). Inline with RSS analysis, yaw right and left can be distinguished clearly via the Doppler shift information in the Doppler spread graphs for 45° (Fig. 4c and 4h) and 90° (Fig. 4d and 4i).

Using equation 3 and assuming that the amplitude of $2V/\lambda$ term is the same for all yaw movements, the range of variation of $\cos\delta \cdot \cos(\beta/2)$ for both yaw right and left were compared for 0°, 45° and 90° configurations. It was observed that while the range of variation of $\cos\delta \cdot \cos(\beta/2)$ for yaw right and left in the 90° configuration has no overlap, the 45° and 0° configurations display overlap and $f_B = 0$ in 180° configuration.

III. METHODOLOGY

The measurements were conducted at 30 GHz in a $3 \times 2.78 \times 5$ m anechoic chamber in the Radio Science Lab at the University of British Columbia. At 30GHz the λ is 1cm, which is much smaller than a proper manually measured head movement of approximately 10-15cm, therefore it is possible to capture it with high accuracy. The reason for using specifically 30GHz frequency is that it met our

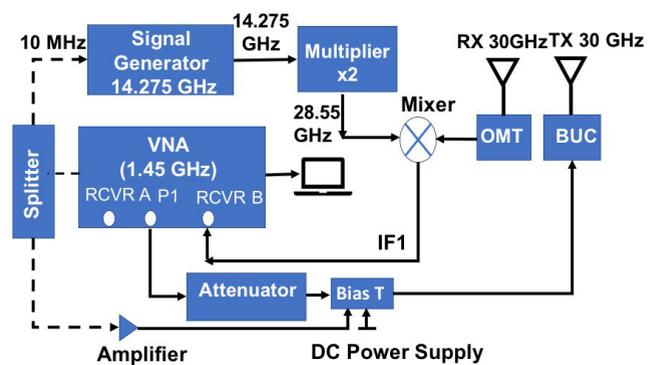


FIGURE 5. System block diagram.

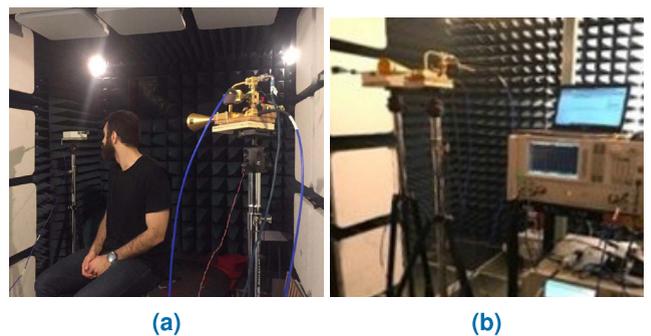


FIGURE 6. The experimental testbed in (a) bistatic configuration, and (b) monostatic configuration.

criteria of separating 3D head turns and it is representative of unlicensed 60GHz frequency, with the change in wavelength only by a factor of 2 ($\lambda = 0.5\text{cm}$) and low attenuation level ($<1.2\text{dB}/10\text{m}$) behaviour much similar to 30GHz. There were complete equipment facilities in our lab for 30GHz frequency, to use them for experiment set up and evaluate our technique. Examining a range of higher frequencies is one of our future work. The chamber provided a static environment with stable positioning of the objects, a certain signal-to-noise ratio, and known clutter, which should afford isolation of signal variation due solely to the human body. Wireless channel impairments are expected in realistic environments and will need to be resolved by future work. The schematic

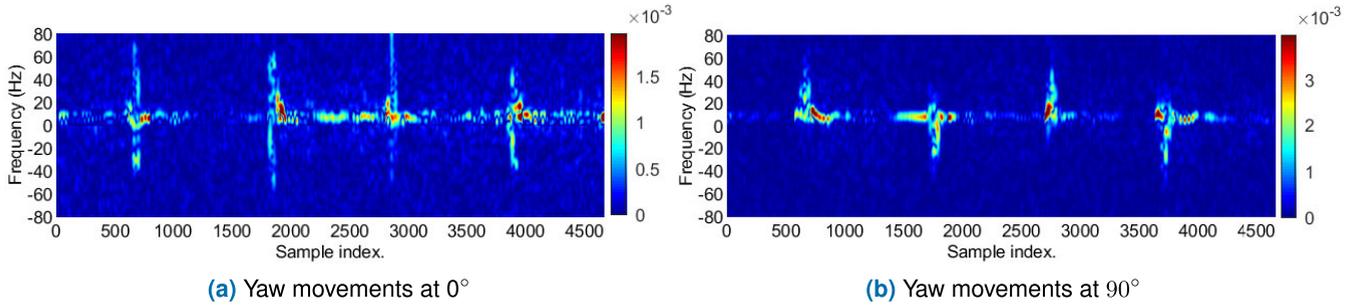


FIGURE 7. Spectrograms for yaw left and right in 0° and 90° configurations.

block diagram and the experimental testbed used in our study are shown in Fig. 5 and Fig. 6.

An E8362C PNA microwave network analyzer was used as both source and VNA, at the TX and RX. For reason of limited VNA frequency range, a block upconverter (BUC) (Norsat 7040STC) was used to upconvert the 1.45 GHz CW source signal to a 30 GHz signal transmitted with a horn antenna (Pasternack PE9850). At the RX side, an orthomode transducer (OMT) (SAT-303-31528-C1-1) was used to separate two orthogonal polarized signals for the RX horn antenna (SAC-2309-315-S2). To monitor the 30 GHz signal received with the VNA, an E8257D analog signal generator created a 14.275 GHz sinusoidal wave, which was doubled with a frequency multiplier and passed through a mixer to downconvert the received signal. A laptop was connected to the VNA via a TCP connection for remote control of experiment operations. The equipment parameters are listed in Table 1.

For all measurement carried out in both monostatic and bistatic configurations, the TX antenna location was fixed at 90 cm away from the human subject. The RX antenna was placed 90 cm away from the human subject at 0°, 45°, 90°, and 180° with respect to the TX antenna, as shown in Fig. 2. The human subject sat in front of the TX antenna such that the head was within the antennas’ aperture, to make sure all head movements were captured. Each type of head movements was performed similarly across all configuration angles, with the human face starting at the TX and ending at an angle to 90° toward or away from the RX for yaw and roll movements. Finally, measurements were collected multiple times for different subject on different times to avoid any environment or subject bias in the results.

A. STFT ANALYSIS OF MOVEMENTS

We performed a discrete-time short-time Fourier transform (STFT) for input signal $x[n]$ with window function $w[n]$. While the FFT of the full t -second dataset captures all Doppler activity in a single spectrum, an STFT is better for identifying the direction, time of occurrence, and duration of movements. STFT takes the Fourier transform of a windowed input signal, with the input signal split into $m = N/\tau$ chunks and the Fourier transform performed for each individually. N is the total number of points in $x(n)$ and τ is the time separation between sections. In this case, time-localized frequency

TABLE 1. Measurement parameters.

Equipment	Parameter	Value
TX antenna	Gain	20 dBi
	Beam width	18.3°
RX antenna	Gain	23 dBi
	Beam width	11°
VNA	Sweep mode	CW
	Number of points	1,201
	Sweep time	5.1 s
	IF BW	10 KHz
	Power	-20 dBm

information can be obtained given by,

$$STFT(x[n]) = \sum_{n=1}^{\infty} x[n]\omega[n - m]e^{-j\omega n}. \tag{4}$$

Windowing function $w(n)$ can be understood as a brick-wall filter of width τ , but it is recommended to use a tapered window, such as a Hann or Hamming function, combined with some overlap between neighboring sections to minimize sidelobe amplitude [28].

Fig. 7 shows the STFT analysis of yaw movements at 0° and 90°. Frequency resolution of the Short Time Fourier Transform (STFT) plot is $\frac{\text{sampling rate}}{\text{FFT size}} = .9141 \text{ Hz/bin}$ where sampling rate = 234 Hz and FFT size = 256. The time resolution is $\frac{\text{window-overlap}}{\text{sampling rate}} = 0.1368\text{s}$ where window = 64 which is a Hamming window and overlap = 32. There are four movements in each spectrogram: yaw right, yaw left, and both repeated. In Fig. 7(a), frequency spreads from negative to positive values in all movement thus, yaw left and right are indistinguishable from each other. In Fig. 7(b), however, each preceding movement has an opposite frequency shift, and makes movements clearly distinguishable.

IV. IMPLEMENTATION

In this section, we describe the steps in our process to detect head movements, refer to Fig. 8. We achieved accurate separation between movements via event boundary detection in which the boundaries were marked by means of an impulsive

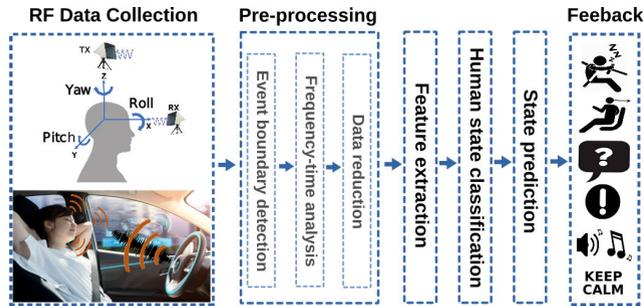


FIGURE 8. 3D head motion detection from wireless signals.

windowing approach and frequency–time STFT analysis was used to calculate the features. For obtaining a distinctive, highly uncorrelated, and easily separable feature set, data reduction was employed and reduced features were extracted accordingly. Then an appropriate machine learning algorithm was applied and trained to predict states of the human head.

A. EVENT BOUNDARY DETECTION

Before one can detect and identify individual activities being performed, windowing the RF signal is required to separate them from each other over time.

Abrupt and instantaneous changes occurring in time-series data from a natural physical environment demands efficient detection of changes, and cost should be optimized. We adopted a concept of finding locations where data values are changing abruptly and utilizing these in marking the boundaries of a window. This is a change-point problem which identifies the points in the input data where statistical attributes fluctuate [29].

In mathematical terms, the input data in ordered sequence can be represented as $y_{1:n} = (y_1, \dots, y_n)$. The output model should include the number of change points, m , along with their locations, $\tau_{1:m} = (\tau_1, \dots, \tau_m)$. The change-point positions must be integers, with each lying in the range between 1 and $n - 1$. We defined $\tau_0 = 0$ and $\tau_m + 1 = n$ and assumed the change points to be ordered such that $\tau_i < \tau_j$ if, and only if, $i < j$. Consequently, the m change points will split the dataset into $m + 1$ segments, with the i th segment containing $y(\tau_{i+1}) : \tau_i$.

To resolve the issues highlighted above, we made use of a maximum-likelihood estimation algorithm that dynamically identifies indices in our data where a significant statistical change has occurred. In particular, we aimed to find all points in our data where the standard deviation has large and abrupt changes.

We refer to these points as the window boundaries, and we assume that one independent activity is being performed between every two adjacent boundaries.

In our 20 seconds of measurement, four distinct head movements were performed, with no movement regions between these. The total number of change points detected is 12–15, with accurate capturing of the head-movement events,

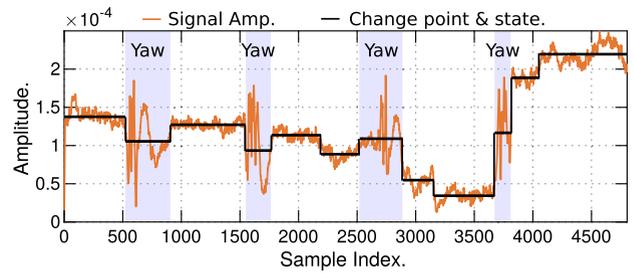


FIGURE 9. Windowing for event detection.

as indicated by Fig. 9. We assigned labels to windows of movements and of no movements accordingly.

B. FREQUENCY–TIME ANALYSIS

1) FEATURE CALCULATION

For each window computed, joint frequency–time analysis was performed and all instantaneous features were obtained from the Doppler spectra. Our pool of extracted features is composed of derived features such as mean, minimum, and maximum values of RSS, alongside frequency, velocity, displacement, and Doppler spread. The pool of multi-dimensional features still requires further processing, to reduce them to a subset that is more robust in distinguishing the set of classes. For this purpose, dimensional reduction was performed.

2) DATA REDUCTION AND FEATURE EXTRACTION

Since the features obtained via the spectrogram analysis are derived features and we allow for a large number of explanatory variables, there is a possibility of high correlations and over-fitting the model, through which some results might not generalize between datasets. To address this problem, the d -dimensional feature set was projected into an l -dimensional feature set, where $l < d$. This method, called principal component analysis (PCA), involves orthogonal transformation of the data matrix, converting the set of correlated variables to linearly uncorrelated variables, referred to as principal components.

The amount of variance in each principal component is explained by its Eigenvalue; PCA is performed by Eigen decomposition of correlation matrix, \mathbf{E}^{PCA} of \mathbf{Z} to transform correlated rows into a new orthogonal coordinate system: $\mathbf{E}^{PCA} = \mathbf{Z}^T \mathbf{Z} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^{-1}$ where $\mathbf{\Lambda}$ is the diagonal matrix of Eigenvalues $\lambda_{(k)}$ of $\mathbf{Z}^T \mathbf{Z}$ and λ_1 has the highest variance [30].

Dimensionality reduction with PCA was a very critical step for our multi-dimensional dataset which helped to remove the correlations between features and to extract the most distinguishable features.

3) CLASSIFICATION AND PREDICTION OF HUMAN STATE

Supervised learning was employed to train the classification model, based on the labeled data, and predict values for an unknown dataset. To achieve this, the data was split into a training set (75%) and a test set (25%), using a

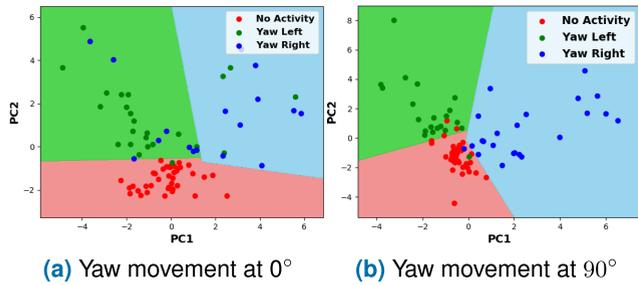


FIGURE 10. Class separation with SVM for 0° and 90° TX-RX configurations.

cross-validation library from Scikit-learn. The classification model was built with the training set and its performance was evaluated with the test set. Our calculated feature set contains data in non-standard form – the independent variables’ values are not on the same scale. Hence, to apply PCA and the classification model, the Scikit-learn pre-processing library was used to scale the feature set. After that, PCA was applied to the standardized training and test set. It was observed that our dataset contains greater than 50% variance in the first two principal components. Since it is still possible to visualize the separation of boundaries, as shown in Fig. 10, we opted to retain the first two components.

A support-vector machine (SVM) was used for classification and a discriminating or conditional classifier was designed to create a hyperplane that separates between classes. It is a supervised learning method that takes labeled training data of n points in the form $(\vec{x}_1, y_1), \dots, (\vec{x}_n, y_n)$ as input and determines the optimal hyperplane to categorize new classes. In two-dimensional space, where the y_i values are either 1 or -1 , the hyperplane is a single line dividing the plane into two parts, to categorize the data into two classes [31]. SVM’s are reliable classical machine learning classifiers used in wireless sensing for applications such as human fall detection [32], [33].

V. EVALUATION AND RESULTS

The following metrics that are commonly used in machine learning algorithms were applied to evaluate our algorithm’s performance,

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \quad (5)$$

$$Precision = TP/(TP + FP) \quad (6)$$

$$Recall = TP/(TP + FN), \quad (7)$$

where TP = true positives, FP = false positives, TN = true negatives, and FN = false negatives. To clarify these metrics, the example of the “yaw movement” class is considered here. TP would correctly detect yaw movements and FP indicates non-yaw movements detected as yaw. TN indicates non-yaw movements correctly detected as non-yaw, and FN presents actual yaw movement detected as non-yaw. Recall and precision values provide more concrete evaluation of the detection algorithm. Recall, or sensitivity, is the percentage of actual yaw movements correctly detected as yaw by the

system, and precision, or positive predictive value, is the proportion of yaw-movement detection that reflect actual yaw movements. F1 score is the harmonic mean of precision and recall and accuracy refers to the fraction of correctly detected head movement classes out of all the events detected. One should strive for algorithm performance levels of 100% for accuracy, precision, recall, and F1.

Extensive research was performed to achieve maximum accuracy for detecting translational and rotational movements independently and in combination. The former refers to the system being able to detect the given type of movement, translation or rotation, while considering all other movements to be noise. Combined movement detection refers to the system’s detection of both translational and rotational movements.

Fig. 11 shows the confusion matrices for translational and rotational movements in monostatic and bistatic configurations. From RSS and Doppler spread empirical study, it was inferred that bistatic angles of 45° and 90° preserve the most relevant information about rotational movements. Further analysis from STFT features for these two configurations’ data revealed reaching maximum accuracy at 90° for complete 90° head rotation. For pitch movements, as shown in Fig. 11(a) and (b), 0° configuration shows a maximum accuracy of 96%, precision and recall of 90%, while the accuracy falls to 85%, precision to 73% and recall to 56% with 90° configuration. For yaw movements (Fig. 11(c) and (d)), in contrast, the 90° configuration performs best, with an accuracy level of 91% with precision and recall value of 96%. In a monostatic configuration, the accuracy drops to 77% with precision of 53% and recall of 57%.

Fig. 12 shows the confusion matrices for combined translational and rotational movements. For combined pitch and yaw, the maximum 80% accuracy was obtained in the monostatic configuration, while the bistatic 90° configuration yielded an accuracy of 76%, refer to Fig. 12(a) and (b), respectively. This shows that we cannot accurately detect complete 3D movement with one type of configuration. It is predicted strongly that both monostatic and bistatic configurations are needed simultaneously for detecting combined translation and rotation movements, though due to hardware limitations, it was skipped in our measurements. Instead, we evaluated the data of combined 0° and 90° configurations in Fig. 12 which show that the accuracy improves to 88%, with precision and recall of 90%, for combined movements’ detection. Roll is another translational movement that behave similar to the pitch movement with a greater Doppler impact at 90°, due to the direction of movement, than at 0° configuration. Performing analysis with a feature set containing labels for all pitch, roll, and yaw, 84% accuracy was reached when pitch movements were captured in a 0° and roll and yaw movements in a 90° configuration.

The evaluation metrics calculated for head movements are in line with our claim that a bistatic configuration significantly improves the detection accuracy for rotational movements while translational movements are handled best by

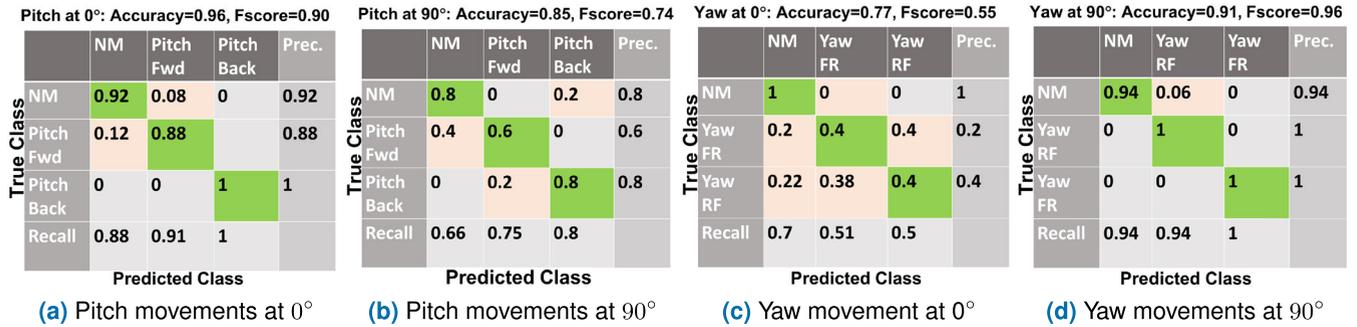


FIGURE 11. Confusion matrices for pitch and yaw movements in 0° and 90° configurations. Accuracy, precision and recall are calculated using Equation 5, 6 and 7. Classes NM, PitchFwd, PitchBack, YawFR, YawRF, represent no movement, pitch forward, pitch backward, yaw right and yaw left respectively.

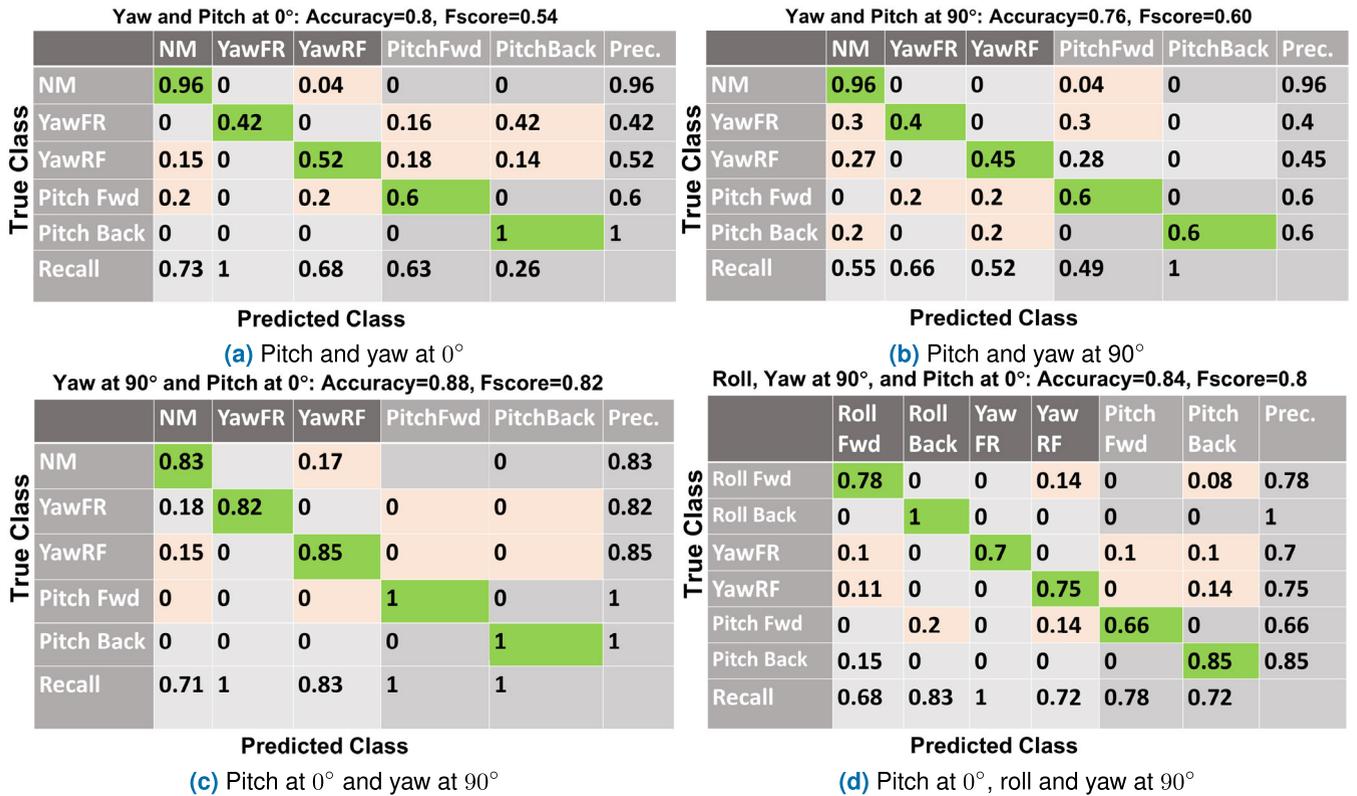


FIGURE 12. Confusion matrices for combined yaw and pitch movements at (a) 0° configuration, (b) 90° configuration, and (c) combined configuration at 0° and 90°, and (d) pitch movement at 0° and roll and yaw movements at 90°.

a monostatic configuration. From our individual movement detection results, we strongly believe that the accuracy will be significantly improved if data is collected in a multistatic configuration.

VI. LIMITATIONS

The primary focus of this research is to solve a complex problem of separating rotational and translational movements from each other with high accuracy. The technique is demonstrated for a single person as there are potential solutions to remove the impact of another person, for instance one sitting behind. One way is to utilize the direct relationship of RMS delay to transmitted power. In both monostatic and bistatic configurations, when a passenger is behind the

driver, the driver blocks the transmitted signal, thus less power is received by the passenger. In case of any reflection from the passenger, the echo will have a higher RMS delay spread especially if it goes through multiple reflections which implies less received power compared to the reflected signal from the driver. At the receiver, any received signal below a certain threshold can be simply ignored.

The other way in which this challenge can be addressed with our solution is the utilization of appropriate antenna full beam width in bistatic configuration to minimize the reflection from movements outside the beam width.

In order to enhance the robustness of the approach we need to counter for any changes in the accuracy due to adverse road conditions. The current results are achieved from

measurements performed in an anechoic chamber to primarily separate the translational and rotational head movements. This challenge can be resolved by incorporating car movement information from existing motion sensors in modern cars. These motion sensors are with integrated three-axis accelerometer, three-axis gyroscope and three-axis magnetometer which provide accurate information about the car movement process, any sudden changes in velocity, angle or direction. The data collected by motion sensors for usual and unusual road conditions could be utilized to spot any abrupt changes in the wireless signals. The change in the signal behaviour could be treated as car movement noise and filtered out to prevent the changes in the accuracy of human movement detection.

VII. CONCLUSION AND FUTURE WORK

Human head tracking is an intriguing concept especially in ADAS, partially and conditionally autonomous driving due to its potential in monitoring human behavior. Research into RF sensing is yielding effective tools which, in combination with the availability of cheaper high-frequency devices, allow for high-resolution movement detection. In this paper, to achieve in-vehicle 3D head tracking, a new technique was proposed that exploited a 30 GHz millimeter-wave Doppler radar with a car-like arrangement within an anechoic chamber. To detect rotational vs. translational movements which has not been studied yet to the best of our knowledge, the usage of bistatic and monostatic geometry for antennas was evaluated. Data reduction and joint time–frequency analysis provided a distinctive feature set to isolate dynamic rotational and translational head movements via class memberships.

From our experiments and analysis of translational and rotational movements at 0° , 45° , 90° , and 180° configurations, we demonstrated the advantages of a bistatic over a monostatic configuration for RF sensing of rotational movements. It was shown that 90° is the optimum bistatic angle to capture the full rotational (yaw) movement with 92% accuracy. Our combined approach of refined configuration and frequency–time analysis shows that translation along the radar line of sight is more easily distinguished in the monostatic configuration and that detecting all classes of rotational and translational movement within one system requires additional RX units at, 0° and 90° , for an accuracy above 88%.

Our approach can be characterized as comprising marking boundaries to detect individual movements and separate them from each other, after which the pool of features is computed via STFT analysis, where a distinctive feature set for dynamic movements' separation is obtained by removing the correlations between features via a PCA-based feature extraction technique. A supervised machine learning classification algorithm SVM was used to train our model and predict the unknown movement classes. This non-invasive and simpler approach to head tracking holds potential to improve driver state monitoring in ADAS and conditionally autonomous vehicles and help intelligent car assistants guarantee a smooth and safe journey. In particular, it is vital to predict the driver's

capacity to supervise or maneuver the vehicle in case of unexpected road events.

The authors intend to perform these experiments in real-world vehicle conditions to verify our approach's practicality. On account of the beamforming capability of mmWave signal and the static nature of the in-car environment, we expect to see similar accuracy with regard to head movements. Also, we plan to extend our system's movement-detection capability and classes by detecting small head displacements and conducting comparisons with fine-motion-tracking devices.

ACKNOWLEDGMENT

The authors would like to thank A. Hughes, E. Xu, and P. Zarei for their meaningful contributions to carrying out this research.

REFERENCES

- [1] A. Tawari and M. M. Trivedi, "Robust and continuous estimation of driver gaze zone by dynamic analysis of multiple face videos," in *Proc. IEEE Intell. Vehicles Symp. Process.*, Jun. 2014, pp. 344–349.
- [2] C. Karatas, L. Liu, M. Gruteser, and R. Howard, "Single-Sensor Motion and Orientation Tracking in a Moving Vehicle," in *Proc. 15th Annu. IEEE Int. Conf. Sens., Commun., Netw. (SECON)*, Jun. 2018, pp. 1–9.
- [3] P. Smith, M. Shah, and N. da Vitoria Lobo, "Monitoring head/eye motion for driver alertness with one camera," in *Proc. 15th Int. Conf. Pattern Recognit. (ICPR)*, Nov. 2002.
- [4] E. Murphy-Chutorian and M. M. Trivedi, "Head pose estimation and augmented reality tracking: An integrated system and evaluation for monitoring driver awareness," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 2, pp. 300–311, Jun. 2010.
- [5] Y. Dong, Z. Hu, K. Uchimura, and N. Murayama, "Driver inattention monitoring system for intelligent vehicles: A review," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 2, pp. 596–614, Jun. 2011.
- [6] Y. Xing, C. Lv, Z. Zhang, H. Wang, X. Na, D. Cao, E. Velenis, and F.-Y. Wang, "Identification and analysis of driver postures for in-vehicle driving activities and secondary tasks recognition," *IEEE Trans. Comput. Soc. Syst.*, vol. 5, no. 1, pp. 95–108, Mar. 2018.
- [7] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-home gesture recognition using wireless signals," in *Proc. 19th Annu. Int. Conf. Mobile Comput. Netw.*, New York, NY, USA: ACM, 2013, pp. 27–38.
- [8] K. J. Lee, C. Park, and B. Lee, "Tracking driver's heart rate by continuous-wave Doppler radar," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2016, pp. 5417–5420.
- [9] Z. Yang, M. Bocca, V. Jain, and P. Mohapatra, "Contactless breathing rate monitoring in vehicle using UWB radar," in *Proc. 7th Int. Workshop Real-World Embedded Wireless Syst. Netw.*, 2018, pp. 13–18.
- [10] M. Raja, V. Ghaderi, and S. Sigg, "WiBot! In-vehicle behaviour and gesture recognition using wireless network edge," in *Proc. IEEE 38th Int. Conf. Distrib. Comput. Syst. (ICDCS)*, Jul. 2018, pp. 376–387.
- [11] F. Wang, J. Liu, and W. Gong, "WiCAR: Wifi-based in-car activity recognition with multi-adversarial domain adaptation," in *Proc. Int. Symp. Quality Service*, p. 19, 2019.
- [12] X. Xie, K. G. Shin, H. Yousefi, and S. He, "Wireless CSI-based head tracking in the driver seat," in *Proc. 14th Int. Conf. Emerg. Netw. Exp. Technol. (CoNEXT)*, 2018, pp. 112–125.
- [13] M. Raja, A. Hughes, Y. Xu, P. Zarei, D. G. Michelson, and S. Sigg, "Wireless multifrequency feature set to simplify human 3-D pose estimation," *IEEE Antennas Wireless Propag. Lett.*, vol. 18, no. 5, pp. 876–880, May 2019.
- [14] J. Xiao, Z. Zhou, Y. Yi, and L. M. Ni, "A survey on wireless indoor localization from the device perspective," *CSURACM Comput. Surv.*, vol. 49, no. 2, pp. 1–31, Jun. 2016.
- [15] I. Teyeb, O. Jemai, M. Zaied, and C. B. Amar, "Towards a smart car seat design for drowsiness detection based on pressure distribution of the driver's body," in *Proc. Int. Conf. Softw. Eng. Adv.*, 2016, pp. 217–222.
- [16] J. Choi, V. Va, N. Gonzalez-Prelcic, R. Daniels, C. R. Bhat, and R. W. Heath, Jr., "Millimeter-wave vehicular communication to support massive automotive sensing," *IEEE Commun. Mag.*, vol. 54, no. 12, pp. 160–167, Dec. 2016.

- [17] T. S. Rappaport, J. N. Murdock, and F. Gutierrez, "State of the art in 60-GHz integrated circuits and systems for wireless communications," *Proc. IEEE*, vol. 99, no. 8, pp. 1390–1436, Aug. 2011.
- [18] C. N. Prudente, R. Stilla, C. M. Buetefisch, S. Singh, E. J. Hess, X. Hu, K. Sathian, and H. A. Jinnah, "Neural substrates for head movements in humans: A functional magnetic resonance imaging study," *J. Neurosci.*, vol. 35, no. 24, pp. 9163–9172, Jun. 2015.
- [19] R. J. Leigh, *The Neurology of Eye Movements* (Contemporary Neurology Series) 5th ed. Oxford, U.K.: Oxford Univ. Press, 2015.
- [20] K. Friedrich, "Bistatic radar: Principles and practice," *Bull. Amer. Meteorol. Soc.*, vol. 89, no. 12, pp. 19–23, Dec. 2008.
- [21] C. J. Baker and H. D. Griffiths, *Bistatic and Multistatic Radar Sensors for Homeland Security*, vol. 2. Dordrecht, The Netherlands: Springer, 2006.
- [22] L. M. H. Ulander and T. Martin, "Bistatic ultra-wideband SAR for imaging of ground targets under foliage," in *Proc. IEEE Int. Radar Conf.*, Jun. 2005, pp. 419–423.
- [23] M. A. Richards, W. A. Holm, and J. Scheer, *Principles Modern Radar: Basic Principles* (Institution of Engineering and Technology). New York, NY, USA: SciTech, 2010.
- [24] D. Ao, Y. Li, C. Hu, and W. Tian, "Accurate analysis of target characteristic in bistatic SAR images: A dihedral corner reflectors case," *Sensors*, vol. 18, no. 2, p. 24, Dec. 2017.
- [25] F. Colone, D. W. O'Hagan, P. Lombardo, and C. J. Baker, "A multistage processing algorithm for disturbance removal and target detection in passive bistatic radar," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 45, no. 2, pp. 698–722, Apr. 2009.
- [26] N. J. Willis, *Bistatic Radar*, vol. 2. Rijeka, Croatia: SciTech, 2005.
- [27] R. Fu, Y. Ye, N. Yang, and K. Pahlavan, "Doppler spread analysis of human motions for Body Area Network applications," in *Proc. IEEE 22nd Int. Symp. Pers., Indoor Mobile Radio Commun.*, Sep. 2011, pp. 2209–2213.
- [28] *Instrument Fundamentals*, Nat. Instrum., Austin, TX, USA, 2017. [Online]. Available: <https://www.ni.com/en-ca/innovations/white-papers/06/understanding-ffts-and-windowing.html>
- [29] R. Killick, P. Fearnhead, and I. A. Eckley, "Optimal detection of change-points with a linear computational cost," *J. Amer. Stat. Assoc.*, vol. 107, no. 500, pp. 1590–1598, Dec. 2012.
- [30] S. Palipana, D. Rojas, P. Agrawal, and D. Pesch, "FallDeFi: Ubiquitous fall detection using commodity Wi-Fi devices," *ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, no. 4, pp. 1–25, Jan. 2018.
- [31] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
- [32] H. Wang, D. Zhang, Y. Wang, J. Ma, Y. Wang, and S. Li, "RT-fall: A real-time and contactless fall detection system with commodity WiFi devices," *IEEE Trans. Mobile Comput.*, vol. 16, no. 2, pp. 511–526, Feb. 2017.
- [33] D. Zhang, H. Wang, Y. Wang, and J. Ma, "Anti-fall: A non-intrusive and real-time fall detector leveraging csi from commodity wifi devices," in *Inclusive Smart Cities e-Health*, A. Geissbühler, J. Demongeot, M. Mokhtari, B. Abdulrazak, and H. Aloulou, Eds. Cham, Switzerland: Springer, 2015, pp. 181–193.



MUNEEDA RAJA (Member, IEEE) received the B.Sc. degree in computer software engineering from and NUST, Pakistan, in 2009, and the M.Sc. degree in geoinformatics from Aalto University, Finland, in 2015, where she is currently pursuing the Ph.D. degree.

She was a visiting Ph.D. student with the University of British Columbia, Vancouver, BC, Canada. Her research includes RF-based device-free activity recognition, in particular sentiment-indicators. She aims to exploit the capabilities of RF-signals, to detect and learn human behavior, emotions and attention levels using signal processing, and machine learning algorithms.



ZAHRA VALI received the B.Sc., M.Sc. (Hons.), and Ph.D. (Hons.) degrees in electrical engineering-communications from the Isfahan University of Technology, Isfahan, Iran, in 2009, 2011, and 2018, respectively.

She was accomplished her Ph.D. sabbatical with the University of British Columbia, in 2016. She has also been conducted research with Northumbria University, Newcastle upon Tyne, U.K., since 2013. She has been a Postdoctoral Fellow with the

Radio Science Laboratory, Electrical and Computer Engineering Department, the University of British Columbia, Canada, since December 2018, working in the area of radar target detection. She has research experiences in a variety of academic research projects, e.g., radar and satellite communications, optical wireless communications, underwater communications, and wireless channel modeling.



SAMEERA PALIPANA received the B.Sc. degree (Hons.) in electronics and telecommunication engineering from the University of Moratuwa, Sri Lanka, in 2010, the master's degree from the Department of Information and Communication Technology, University of Bremen, Germany, in 2014, and the Ph.D. degree from the Cork Institute of Technology, Ireland, in 2019. He is currently a Postdoctoral Researcher with the Ambient Intelligence Group, Department of Communications and Networking, Aalto University. His research interests are in device-free-based localization, activity recognition, signal processing, and machine learning using wireless signals.



DAVID G. MICHELSON (Senior Member, IEEE) is currently an Associate Professor and leads the Radio Science Laboratory, Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, Canada. His current research interests include development of propagation and channel models for 5G mm-wave communications, and classification and interpretation of imagery generated by spaceborne synthetic aperture radars. He serves as the Chair of the Canadian National Committee, International Union of Radio Science, from 2018 to 2020, and Canadian representative to Commission F-Wave Propagation and Remote Sensing, from 2015 to 2020. In Fall 2018, he has served as a Guest Researcher with the Wireless Network Division, National Institute of Standards and Technology.



STEPHAN SIGG (Member, IEEE) is currently an Associate Professor with the Department of Communications and Networking, Aalto University. His research interests include the design, analysis and optimization of algorithms for distributed, and ubiquitous systems. Especially, his work covers proactive computing, distributed adaptive beamforming, context-based secure key generation, and device-free passive activity recognition. He is also an Editorial Board Member of the *Journal on Computer Communications* (Elsevier). He has been a Guest Editor of the *Personal and Ubiquitous Computing Systems Journal* (Springer). He has served on the organizing and technical committees numerous prestigious conferences, including the IEEE PerCom and ACM Ubicomp.

...