ASL Recognition Based on Kinematics Derived from a Multi-Frequency RF Sensor Network

Sevgi Z. Gurbuz*, Ali C. Gurbuz†, Evie A. Malaia‡, Darrin J. Griffin§, Chris Crawford¶,
Emre Kurtoglu*, M. Mahbubur Rahman*, Ridvan Aksu*, Robiulhossain Mdrafi†
*Dept. of Electrical and Computer Engineering, University of Alabama, Tuscaloosa, AL
†Dept. of Electrical and Computer Engineering, Mississippi State University, Starkville, MS
‡Dept. of Communicative Disorders, University of Alabama, Tuscaloosa, AL
§Dept. of Communication Studies, University of Alabama, Tuscaloosa, AL
¶Dept. of Computer Science, University of Alabama, Tuscaloosa, AL

***Abstract*—As a means for leveraging technology in the design of Deaf spaces, this paper presents initial results on American Sign Language (ASL) recognition using RF sensing. RF sensors are non-contact, non-invasive, and protective of privacy, making them of special interest for use in personal areas. Using just the kinematic properties of signing as captured by the micro-Doppler signatures of a multi-frequency RF sensor network, this paper shows that native and imitation signing can be differentiated with %99 accuracy, while up to 20 ASL signs are recognized with an accuracy of %72 or higher.**

***Index Terms*—American sign language, gesture recognition, radar micro-Doppler, RF sensing**

I. Introduction

Most research in technologies for the Deaf community have focused on translation using either video or wearable devices. Sensor-augmented gloves [1], [2] have been reported to yield higher gesture recognition rates than camera-based systems [3], [4]; however, they cannot capture information expressed through head and body movement. Gloves are also intrusive and inhibit users in their pursuit of normal daily life, while cameras can raise concerns over privacy and are ineffective in the dark. In contrast, RF sensors are non-contact, non-invasive and do not reveal private information even if hacked.

Although RF sensors are unable to measure facial expressions or hand shapes, which would be required for complete translation, this paper aims to exploit RF sensors for the design of smart Deaf spaces. In this way, we hope to enable the Deaf community to benefit from advances in technologies that could generate tangible improvements in their quality of life.

More specifically, this paper investigates the recognition of ASL signs based on kinematics only, as perceived by a multi-frequency RF sensor network. In fact, RF sensors can acquire a *unique source of information* that is inaccessible to optical or wearable devices: namely, a visual representation of the kinematic patterns of motion via the micro-Doppler signature [5]. Micro-Doppler refers to frequency modulations that appear about the central Doppler shift, which are caused by rotational or vibrational motions that deviate from principle translational motion. In prior work [6], we showed that fractal complexity computed from RF data could be used to discriminate signing from daily activities and that RF data could reveal linguistic properties, such as coarticulation.

This paper shows RF data can be used to distinguish imitation from native signing and up to 20 ASL signs can be recognized with %72 accuracy or greater. In Section II, the experimental datasets acquired are presented. Section III addresses the differences between imitation and native signing, while Section IV presents the results for classification of ASL signs. Finally, in Section IV, conclusions and future work are discussed.

II. RF ASL Datasets

In many ASL recognition studies, non-native signers, who may not know any ASL, are used as an expeditious source of data. However, imitation signing by non-signers results in motor production that does not approximate native sign language production in the speed and stability of motion signatures [7], but has severe spatiotemporal distortion and linguistic errors [8]. It can take learners of sign language at least 3 years to produce signs in a manner that is perceived as fluent by native signers [9].

To study the differences between imitation and native signing, two distinct datasets were acquired: 1) native ASL data from Deaf participants and 2) imitation data from hearing individuals imitating ASL signs based on copy-signing videos. A total of 10 non-signers and 3 native signers were recruited to participate in the study. A total of 180 native signing samples and 2631 imitation signing samples were acquired. Words were selected from the ASL-LEX database (http://asl-lex.org/), choosing words that are higher frequency, but not phonologically related to ensure a more diverse dataset. A complete listing of the words are given in Table I. In all experiments, participants were asked to begin with their hands placed on their thighs, and to return to this position once done signing.

A. Multi-Frequency RF Sensor Network

RF sensors operating at three different transmit frequencies are considered in this work. The Xethru sensor is a low-power ultra-wide band (UWB) impulse radar with transmit frequencies between 7.25 - 10.2 GHz. The range resolution of an RF sensor isgiven by $c/2\beta$, where c is the speed of light and β is the bandwidth. Thus, the Xethru sensor has about 5 cm range resolution. Frequency modulated continuous wave (FMCW) radars at 24 GHz and 77 GHz were also deployed. A 24 GHz Ancortek SDR was operated with bandwidth of 1.5 GHz, while the 77 GHz Texas Instruments device transmitted with a bandwidth of 750 MHz. This results in range resolutions of 10 cm and 20 cm, respectively.

Participants were asked to sit on a bar stool facing three RF sensors (Xethru, 24 GHz and 77 GHz), while two other Xethru sensors were placed off to the side and at a 45° angle elevated 0.9 meters off of the ground. Prompts indicating the sign or sequence of signs to be observed were communicated using a computer monitor placed directly behind the sensor, so that the visual cues would ensure the participant remained facing the sensors throughout the experiment.

B. RF Data Processing

Unlike video, radar measurements are not inherently an image, but are actually a time-stream of complex I/Q data from which line-of-sight distance and radial velocity may be computed. To reveal patterns of motion hidden in the amplitude and frequency modulations of the received signal, time-frequency analysis is often employed. The micro-Doppler signature, or spectrogram, is found from the square modulus of the Short-Time Fourier Transform (STFT) of the continuous-time input signal. It reveals the distinct patterns caused by micro-motions, such as hand gestures and human activity.

Prior to computation of the spectrogram, a 4th order high pass filter (HPF) is applied to remove reflections from stationary objects, such as the walls, tables, and chairs. The STFT itself is computed using Hanning windows with 50% overlap to reduce sidelobes in the frequency domain and convert the 1D complex time stream into a 2D μD signature. Sensor noise and artifacts were mitigated using an isodata thresholding algorithm [10]. Sample ASL micro-Doppler signatures at each transmit frequency are shown in Figure 1.

INDIVIDUAL ASL WORDS			
1. YOU 2. HELLO 3. WALK 4. DRINK 5. FRIEND	6. KNIFE 7. WELL 8. CAR 9. ENGINEER 10. MOUNTAIN	11. LAWYER 12. HOSPITAL 13. HEALTH 14. EARTHQUAKE 15. BREATHE	16. HELP 17. PUSH 18. GO 19. COME 20. WRITE

TABLE I: Listing of ASL signs collected during experiments

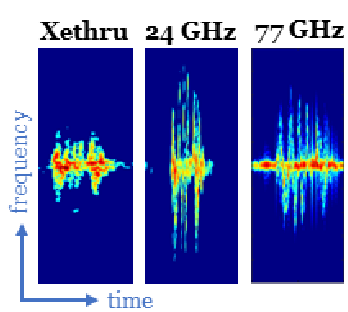

Fig. 1: mD signature of BREATHE for different RF sensors.

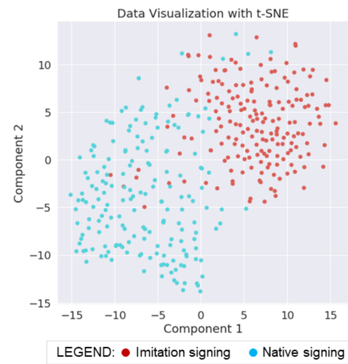

Fig. 2: Comparison of RF data from native ASL users and imitation signers using tSNE.

III. Imitation vs. Native Signing

Differences in signing between native and imitation signers can be revealed through observation of visual data and quantitative analysis of RF ASL data. ASL is a fluid language that minimizes exertion. But imitation signers are often hesitant or awkward, failing to replicate temporal tempo of signing [11], [12]. Other errors of imitation signers include replicating signs with an incorrect number of repetitions, exaggerating movements along inaccurate trajectories, and gross motion errors.

Machine learning can also be used to distinguish between native versus imitation signing. The T-distributed Stochastic Neighbor Embedding (t-SNE) algorithm [13] can be used to visualize the feature space spanned by the data, as shown in Figure 2. To remove any bias due to sample size, 180 samples were randomly selected from the imitation data set, an equal number to the native signing samples utilized. It may be noticed that there is little overlap between imitation and native data in feature space.

In fact, the data may be explicitly discriminated using a Support Vector Machine (SVM) classifier with Radial Basis Function (RBF). The SMOTE [14], [15] algorithm was used to equalize class data by oversampling minority classes with "synthetic" samples. With SVM, the imitation and native datasets were distinguished with %99 accuracy. An important consequence of this result is that the imitation data samples are not suitable for validating true ASL recognition performance, nor are they effective in pre-training models for classification of native data.

IV. Classification of ASL Signs

A wide range of handcrafted features were extracted, after which an optimal subset was selected using the minimum redundancy maximum relevance (mRMR) algorithm [16], [17]. Four types of features were computed: 1) envelope features, frequency warped cepstral coefficients (FWCC), discrete cosine coefficients (DCT), and linear predictive coding (LPC) coefficients. Envelope features have been shown to be significant physical features [18], [19] of the μD signature as

SENSOR / FEATURES	With HPF					NO HPF					FUSION		
	10 GHz Corner	10 GHz Front	10 GHz Side	24 GHz Front	77 GHz Front	10 GHz Corner	10 GHz Front	10 GHz Side	24 GHz Front	77 GHz Front	5 Inputs No HPF	5 Inputs w/HPF	All 10 Inputs
Accuracy	43%	48%	42%	54%	39%	15.6%	29.4%	25.6%	40.6%	52%	**63.6%**	**68.1%**	72.5%
Envelope	6	7	7	3	6	7	6	4	4	6	11	8	8
DCT	22	103	30	24	50	135	15	10	25	60	56	83	86
FWCC	20	134	13	23	42	260	17	4	17	116	45	58	54
LPC	2	6	0	0	2	98	12	2	4	8	8	1	2
Total	50	250	50	50	100	500	50	20	50	190	120	150	150

Fig. 3: Features selected by MRMR algorithm and resulting accuracy with random forest classifier

they describe the outline and peak response of the signature. In this work, the maximum, minimum and mean value of upper and lower envelopes, as well as the difference between the average values of the upper and lower envelope are extracted using the percentile technique [20]. Frequency-warped cepstral coefficients (FWCC) are related to mel-frequency cepstral coefficients, common in speech processing, but differ in that their filter bank is optimized to RF data using genetic algorithms [21]. The Discrete Cosine Transform (DCT) [22] represents a signal as the sum of sinusoidal functions oscillating at different frequencies. Linear predictive coding (LPC) mD [23] computes the coefficients of a forward linear predictor by minimizing the prediction error in a least squares sense.

A total of 932 features are initially extracted for each RF sensor in the network: 7 envelope features, 500 DCT coefficients, 325 FWCC features, and 100 LPC coefficients. For all sensors, features were extracted from the spectrogram both with and without high pass filtering (HPF). This is because while HPF removes stationary clutter, there is also the potential for important low-frequency information to be lost. Next, the mRMR algorithm was applied to select an optimal subset of features from each sensor. The number of features selected was varied between 20 and 250, while four different classifiers considered to evaluate performance: support vector machines (SVM), k-nearest neighbors (kNN), linear discriminant analysis (LDA) using random subspace gradient boosting, and random forest classifier (RFC). During classification, %75 of the data was used for training, while %25 was used for testing. The random forest classifier was found to give the best result.

Table 3 summarizes the features selected by mRMR. Notice that frequency-dependent features, e.g. DCT and FWCC, are heavily favored, followed by envelope features, which capture the extremety of the motion. However, LPC appears to not be very effective for ASL recognition, with very few LPC coefficients being selected.

Moreover, better performance is achieved at the higher transmit frequencies, with the accuracy of the 24 GHz sensor closely followed by that of the 77 GHz sensor. Notice that there is a great benefit to not using a HPF on 77 GHz data, for which the accuracy without the HPF is increased by 13% whereas other sensors benefit from the filtering. Also, the ASL recognition accuracy can be greatly increased by fusing features form all inputs, and performing a feature selection

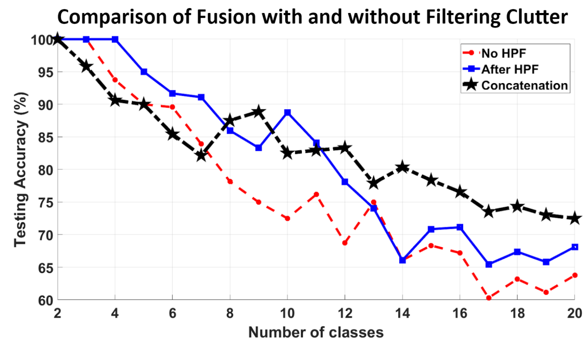

Fig. 4: Classification accuracy versus classes with and without ground clutter removal.

with mRMR. This yields a classification accuracy of 72% for 20 ASL classes - about a 15% - 30% improvement over the results obtained with just a single sensor. A classification accuracy of 95% is attained for recognition of 5 ASL signs. The impact of HPF is compared in Figure 4 as a function of the number of classes.

V. Conclusion

This paper presents initial work on ASL recognition with a multi-frequency RF sensor network, which extracts only kinematic properties of signing. Results show that RF sensors can be used to differential whether the signer is a native signer or a non-signer doing copysigning. This shows the need to validate machine learning algorithms on native signer data. Frequency warped cepstral coefficients (FWCC) are optimized for ASL using genetic algorithms, and in conjunction with Discrete Cosine Transform (DCT) coefficients, and envelope features, used to classify up to 20 ASL signs. Using the minimum redundancy maximum relevance (mRMR) algorithm, an optimal subset of 150 features are selected and input to a random forest classifier to achieve %95 recognition accuracy for 5 signs and %72 accuracy for 20 signs.

VI. Acknowledgements

This work was funded by the National Science Foundation (NSF) CPS Program Award #1932547, #1931861 and NSF NCS Program Award #1734938. Human studies research was conducted under the University of Alabama Institutional Review Board (IRB) Protocol #18-06-1271.

References

[1] N. Tubaiz, T. Shanableh, and K. Assaleh, "Glove-based continuous arabic sign language recognition in user-dependent mode," *IEEE Transactions on Human-Machine Systems*, vol. PP, 03 2015.

[2] B. G. Lee and S. M. Lee, "Smart wearable hand device for sign language interpretation system with sensors fusion," *IEEE Sensors Journal*, vol. 18, no. 3, pp. 1224–1232, 2018.

[3] G. García-Bautista, F. Trujillo-Romero, and S. O. Caballero-Morales, "Mexican sign language recognition using kinect and data time warping algorithm," in *2017 International Conference on Electronics, Communications and Computers (CONIELECOMP)*, 2017, pp. 1–5.

[4] V. Bheda and D. Radpour, "Using deep convolutional networks for gesture recognition in american sign language," *CoRR*, vol. abs/1710.06836, 2017. [Online]. Available: http://arxiv.org/abs/1710.06836

[5] V. Chen, *The Micro-Doppler Effect in Radar*. Artech House, 2019.

[6] S. Z. Gurbuz, A. C. Gurbuz, E. A. Malaia, D. J. Griffin, C. Crawford, M. M. Rahman, R. Aksu, E. Kurtoglu, R. Mdrafi, A. Anbuselvam, T. Macks, and E. Ozcelik, "A linguistic perspective on radar micro-doppler analysis of american sign language," in *2020 IEEE International Radar Conference (RADAR)*, 2020, pp. 232–237.

[7] A. I. Hilger, T. M. Loucks, D. Quinto-Pozos, and M. W. Dye, "Second language acquisition across modalities: Production variability in adult l2 learners of american sign language," *Second Language Research*, vol. 31, no. 3, pp. 375–388, 2015.

[8] J. P. Morford, B. Nicodemus, and E. Wilkinson, "Research methods in psycholinguistic investigations of sign language processing," *Research methods in sign language studies: A practical guide*, p. 209, 2015.

[9] J. S. Beal and K. Faniel, "Hearing l2 sign language learners: How do they perform on asl phonological fluency?" *Sign Language Studies*, vol. 19, no. 2, pp. 204–224, 2018.

[10] I. T. Young, J. J. Gerbrands, and L. J. van Vliet, "Fundamentals of image processing." Delft Institute of Technology, 1998.

[11] A. Cull, *Production of Movement in Users of American Sign Language and its Influence on Being Identified as" Non-Native"*. Gallaudet University, 2014.

[12] R. B. Wilbur and E. Malaia, "Contributions of sign language research to gesture understanding: What can multimodal computational systems learn from sign language research," *International journal of semantic computing*, vol. 2, no. 01, pp. 5–19, 2008.

[13] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.

[14] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *J. Artif. Int. Res.*, vol. 16, no. 1, p. 321–357, Jun. 2002.

[15] T. D. Bufler and R. M. Narayanan, "Radar classification of indoor targets using support vector machines," *IET Radar, Sonar Navigation*, vol. 10, no. 8, pp. 1468–1476, 2016.

[16] Hanchuan Peng, Fuhui Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.

[17] B. Tekeli, S. Z. Gurbuz, and M. Yuksel, "Information-theoretic feature selection for human micro-doppler signature classification," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 54, no. 5, pp. 2749–2762, 2016.

[18] S. Z. Gurbuz, B. Erol, B. Cagliyan, and B. Tekeli, "Operational assessment and adaptive selection of micro-doppler features," *IET Radar, Sonar Navigation*, vol. 9, no. 9, pp. 1196–1204, 2015.

[19] Y. Kim, S. Ha, and J. Kwon, "Human detection using doppler radar based on physical characteristics of targets," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 2, pp. 289–293, 2015.

[20] P. V. Dorp and F. C. A. Groen, "Feature-based human motion parameter estimation with radar," *IET Radar, Sonar Navigation*, vol. 2, no. 2, pp. 135–145, 2008.

[21] B. Erol, M. G. Amin, and S. Z. Gurbuz, "Automatic data-driven frequency-warped cepstral feature design for micro-doppler classification," *IEEE Trans. Aero. Elec. Sys.*, vol. 54, no. 4, pp. 1724–1738, 2018.

[22] P. Molchanov, J. Astola, K. Egiazarian, and A. Totsky, "Ground moving target classification by using dct coefficients extracted from micro-doppler radar signatures and artificial neuron network," in *Proc. Microwaves, Radar, and Remote Sensing Symp.*, 2011, pp. 173–176.

[23] R. J. Javier and Y. Kim, "Application of linear predictive coding for human activity classification based on micro-doppler signatures," *IEEE Geosci. Rem. Sens. Lett.*, vol. 11, no. 10, pp. 1831–1834, 2014.