A Linguistic Perspective on Radar Micro-Doppler Analysis of American Sign Language

Sevgi Z. Gurbuz*, Ali C. Gurbuz†, Evie A. Malaia‡, Darrin J. Griffin§, Chris Crawford¶,
M. Mahbubur Rahman*, Ridvan Aksu*, Emre Kurtoglu*, Robiulhossain Mdrafi†,
Ajaymehul Anbuselvam¶, Trevor Macks*, Engin Ozcelik*
*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*—Although users of American Sign Language (ASL) comprise a significant minority in the U.S. and Canada, people in the Deaf community have been unable to benefit from many new technologies, which depend upon vocalized speech,and are designed for hearing individuals. While video has led to tremendous advances in ASL recognition, concerns over invasion of privacy have limited its use for in-home smart environments. This work presents initial work on the use of RF sensors, which can protect user privacy, for the purpose of ASL recognition. The new offerings of 2D/3D RF data representations and optical flow are presented. The fractal complexity of ASL is shown to be greater than that of daily activities - a relationship consistent with linguistic analysis conducted using video.**

***Index Terms*—American Sign Language, radar micro-Doppler, RF sensing, gesture recognition**

I. Introduction

Users of American Sign Language (ASL) make up over 1 million people in the U.S. and Canada, based on statistics provided by Gallaudet University (the world's only university designed to be barrier-free for deaf and hard of hearing students located in Washington, D.C.). People in the Deaf community, who rely on ASL as their primary mode of communication, rely heavily on technology as an assistive device as they navigate communication/language barriers that status quo society often creates. Unfortunately, many technologies are designed for hearing individuals, where vocalized speech is the preferred mode of communication, and has driven a burgeoning market of voice recognition software and voice-controlled devices. This precludes the Deaf community from benefiting from advances in technology, which, if designed to be compatible with ASL, could in fact generate tangible improvements in their quality of life.

Research related to technologies for the deaf or hard of hearing (HoH) has been ongoing for the past three decades, but, has primarily focused on camera-based and wearable technologies, such as gloves or wrist bands containing accelerometers and other sensors to translate sign language into voice or text. Among these approaches, sensor-augmented gloves have been reported to typically yield higher gesture recognition rates than camera-based systems. However, such wearable gloves cannot capture intricacies of sign languages offered through head and body movements [1], [2]. This issue is addressed by optical sensors, however, video cameras trigger user concerns over privacy and require light to be effective.

Moreover, previous investigations of these existing prototypes often fail to involve participants and investigators fluent in ASL. In a preliminary focus group we conducted, Deaf participants reacted negatively to the idea of having to use anything wearable in their daily lives. Wearable technology limits signer's freedom in conducting daily activities and is not designed with ASL movements and language constraints in mind. In contrast, the Deaf participants reported that while they used on a regular basis some form of video-based technology for communication in their jobs; those video technologies have limitations (such as a narrow field-of-view, privacy issues, and reliannce on light). Although video tends to be viewed favorably for interpersonal communication with society-at-large, Deaf participants in our pilot study focus group lamented its limitations – video usage was dependent upon being in an office/work environment or with access to cell phones (and battery life). A significant thematic point of discussion that occurred with the Deaf participants was concern over technology enabling invasion of privacy and potential surveillance of their personal and private lives.

In contrast, especially in the context of technologies for the deaf, RF sensors have several advantages over alternative sensing modalities, which make them uniquely desirable. RF sensors are non-contact and completely private, fully operational in the dark, and can even be used for through-the-wall sensing. Most importantly, RF sensors can acquire a *new source of information* that is inaccessible to optical sensors: visual representation of kinematic patterns of motion via the micro-Doppler signature [3], as well as more accurate velocity measurements and range profiles.

This paper represents the first publication investigating RF sensor-based ASL recognition and offers a linguistic perspective to radar signal understanding. In Section II, we first discuss the importance and contribution of kinematics to the recognition of ASL. Next, in Section III, the radar studies conducted with Deaf participants is described. In Section IV, the radar measurements acquired for different signs are

Experiment 1: Individuals Words	Experiment 2: Sequences of Sentences
1. YOU	1. MY FRIEND MARRY AGAIN. WIFE FIRST DIE LONG~AGO. NOW HAVE WIFE.
2. HELLO	2. KNOW~THAT (NAME) SICK. DOCTOR ORDER. STAY HOME MUST. LATER GO HOSPITAL. FOR TESTS VARIOUS~THINGS ALL~OVER~BODY.
3. WALK	
4. DRINK	3. CAR MY OLD COLLAPSE, MAN #GARAGE TELL~ME. #FIX THINK DOUBT HE. WELL, HAVE~TO BUY NEW #CAR.
5. FRIEND	
6. KNIFE	4. MY MOTHER OLD SICK. MIND CLEAR FINE, BUT WEAK, CAN'T WALK HERSELF, CAN'T.
7. WELL	
8. CAR	5. YESTERDAY SEE ME FRIEND DAUGHTER. OLD 3 SHE. IMPRESS ME, EXPERT FINGERSPELL. HARD~TO~BELIEVE.
9. ENGINEER	
10. MOUNTAIN	6. PLAN GO #FLA. MONEY PLAN 4 DAYS, EXACT. EVERYTHING~UPSET, CAR COLLAPSE. MUST STAY 8 DAYS. BROKE NOW ME.
11. LAWYER	
12. HOSPITAL	7. PARTY DEAF #CLUB, RESPONSIBILITY BUT NOT MY RESPONSIBILITY. ME GO, LOOK~AROUND. NO CLEAN. ME TELL, MUST HELP. ALL RESPONSIBILITY TO CLEAN.
13. HEALTH	
14. EARTHQUAKE	
15. BREATHE	8. WRONG YOU. MEETING STARTS TIME TWO, YOU~TELL~ME NOON. WHAT'S …. YOU.
16. HELP	
17. PUSH	9. SHOCK ME. DISCOVER GOOD FRIEND DIE. THINK HEART~ATTACK. NOT~KNOW…SEEM SICK HE, NOT~KNOW ME.
18. GO	
19. COME	10. NEWSPAPER READ ME, AWFUL #STORM IN #FLA. HOMES, CARS, TREES DESTROY. PEOPLE ABOUT 25 DIE, ABOUT NOT SURE.
20. WRITE	

Fig. 1: Listing of signs/sign sequences collected during experiments

presented and discussed via consideration of micro-Doppler signatures and range-Doppler videos. The concept of optical flow, which originates in video processing to capture the kinematics of each pixel, is presented in Section V. The optical flow of Kinect measurements as well as radar data is compared for the cases of sign language usage and daily activities. It is observed that, on average, the fractal complexity of the optical flow for daily activities is lower than that of ASL usage. In Section VI, conclusions and future work are discussed.

II. Linguistic Significance of Kinematics

In speech communication, quantitative measurements of the temporal dynamics have resulted in fundamental insights into perceptual and mathematical properties of information exchange [4] . Temporal quantification of the properties of signed languages has been, to date, substantially behind that of speech due to the higher dimensionality of visual modality. When sign language linguistics research began in 1960s, signs had been defined based on their static properties: handshape (HS), place of articulation (i.e. location of the articulator/hand at the beginning and end of the sign), and handshape orientation [5].

An early study of signers' perception of writing in point-light displays [6] has demonstrated that signers viewing the dynamics of hieroglyph writing can tell the difference between 'strokes' (information-bearing portions of point-light movement) and 'transitions' (movement of the point-light from the end of one meaningful portion, to the beginning of another). A 2x2 Latin Square design that assessed the difference in perception between signers and non-signers, and users of Chinese and English, showed that sensitivity to transitions was due entirely to experience with sign language, and not due to experience with hieroglyphic writing systems.

Current neurolinguistic research indicates that dynamic properties of signs (speed and temporal contour of motion) contribute crucial linguistic information to the meaning of signs [7], [8]. Analysis of information content in speech vs. everyday motion using the visual properties of the signal and optical flow [9], [10] has indicated that signers transmit more information (in the sense of mathematical entropy) than humans carrying out dynamic tasks, and that the intelligibility of a signing stream is crucially dependent on the ability to parse entropy changes in visual information [11], [12].

Thus, although radar cannot pick up static hand shapes, it does allow for improved measurements of the temporal dynamics of signs in conjunction with shape dynamics, combining information picked up from the moving hands with the information on other articulators (head and body). This approach unites the data from the better-studies parameters of handshape, orientation, and place of articulation with dynamic data on motion. Our goal in the present analysis is to build on analytical understanding of human signal parsing for sign language [11] and demonstrate that radar data can be used for distinguishing between signing and biological motion (e.g. cooking, laundry-folding) that can go on inside the house, but with better sensitivity than what has been shown for 2D video data [9], [10].

III. Study Data

As an initial pilot study to investigate radar measurements of the linguistic and kinematic properties of ASL, three Deaf participants were recruited. The study began with a 30-minute informative session where researchers described each task. After being informed about the content and aims of the study, participants signed informed consent forms. Study participants were then asked to demonstrate ASL signs during two conditions: 1) individual words and 2) sequences of sentences. The study lasted approximately 75 minutes. Participants received a $25 gift card for completing the study. Figure 1 provides a complete listing of the words and sentences comprising these experiments. Three repetitions of each word were collected per participant, resulting in a total of 9 samples per word. 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. Each sequence of sentences was collected once per participant, resulting in 3 samples per sentence. Sentences were chosen from those used in previous linguistic studies of ASL by one of the co-authors. In all experiments, participants were asked to begin with their hands placed on their thighs, and to return to this position once done signing.

Radar measurements were acquired using a Texas Instruments 77 GHz FMCW transceiver, with a transmit waveform bandwidth of 750 MHz. Participants were asked to sit on a bar stool facing the RF sensor and a Kinect sensor at a distance of roughly 1 meter. 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. The Kinect sensor data have been used for comparison with radar data, and to annotate the micro-Doppler signatures and optical flow plots given in this paper. Annotations were performed by a Child-of-Deaf-Adult (CODA) fluent in ASL who made manual notations of the Kinect video frames, which were then correlated to the time axis of the radar data.

IV. Radar Measurements of ASL

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. In this work, the RF data was visualized in two ways: 1) as a short-time Fourier transform of the data, computed to find the spectrogram, or *micro-Doppler signature* of the motion; and 2) as a *range-Doppler video*, where each frame is a range-Doppler map formed by computing the 2D FFT of the data for each sweep of the FMCW radar [13], [14].

A. Experiment 1: Single Words

Illustrative samples of the micro-Doppler signatures for several ASL signs are given in Figure 2. Due to space constraints, not all signatures could be shown, however, there are a number of observations important to note:

- *The starting position of the hand affects the initial and final frequency components measured.* In the case of our experiments, participants started with hands on the thighs. As signing requires raising the arms, typically with a motion away from the radar, towards the body and at the level of the chest, this motion resulted in negative frequency spikes at the beginning of all samples.
- *For synchronized motions where both hands move in tandem, or one hand is moving alone, the sequential nature of the motion is evident from the micro-Doppler signature.* The signature has either a positive or a negative Doppler frequency at a given time, not both. Examples include the signs for `you` and `health`.
- *In signs where the hands move complementarily, e.g. opposite directions, both positive and negative micro-Doppler components are present simultaneously.* Examples include `walk` and `friend`.
- *When more surface area of the hand(s) faces the radar line-of-sight, the received power of the signal may be observed to be greater.* Examples include `drink` and `breathe`. Notice that the `drink` micro-Doppler signature has two vertical spikes; one due to raising the hand cupped, the second lowering the hand. The first peak has a greater intensity as the outside of the hand faces the radar. Once virtual cup is lifted, the side of the hand faces the radar, and this dominates the return signature, but with lesser intensity.
- *In signs with repetitious movements, the number of repetitions can be counted from the micro-Doppler signature*. Examples include `knife` and `walk`.
- *The affect of aspect angle between the line-of-sight and direction of motion can be observed in the signatures.* For example, the sign for `help` involves primarily vertical movement, which is tangential to the radar line-of-sight, and hence has a low Doppler frequency. As the hand moves higher, however, non-zero Doppler does appear, resulting in vertical streaks in the signature as the hands move upwards and downwards.
- *The effects of occlusion can be observed in some signatures.* Consider the sign for `write`, which is expressed by holding the left hand and dragging the right hand (with fingers as though holding a pen) across the inside of the palm. Whereas the sliding motion in `knife` is clearly visible (fingers graze top of the fingers in left hand), in the sign for `write`, the left hand shields the right hand swiping motion - at least partially - from the radar.

The range-Doppler videos enable the simultaneous observation of the radial distance to the radar as well as the radial velocity. An example showing a sequence of several frames for the sign of `breathe` is given in Figure 3. The participant is initially sitting on a stool with hands on knees. As the hands move upward towards the chest, the peak power is received from the torso, accompanied by a slight negative velocity, due to the motion of the hands away from the radar (Frame 323). Once the hands reach the chest, momentarily there is a pause before the hands move towards the radar. Frame 328 shows the instant after the pause, where still the largest reflection is at 0 Hz from the torso, but now there are some slow positive velocity components. The forward motion reaches its fastest point in Frame 331, and both the torso return as well as the return from the hands and arms (now at a large positive velocity) can be visible from the range-Doppler video. When the extension of the hands is at is farthest from the body (closest to the radar), again there is a moment of zero velocity (Frame 333) followed by a large negative velocity as the hands move back towards the chest (Frame 334).

It may be observed that the hands move about 2 range bins closer to the radar than the torso. At 750 MHz, each range bin corresponds to about 20 cm displacement. For a person with average arm length, during the course of enacting

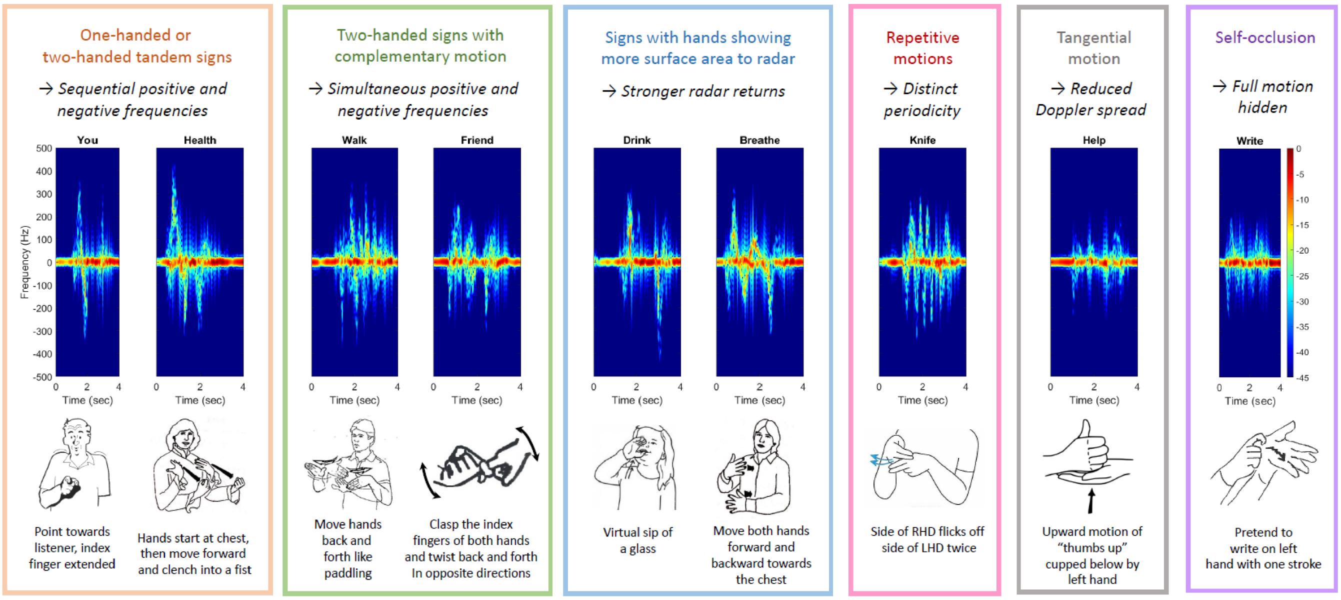

Fig. 2: Micro-Doppler signatures of individual words.

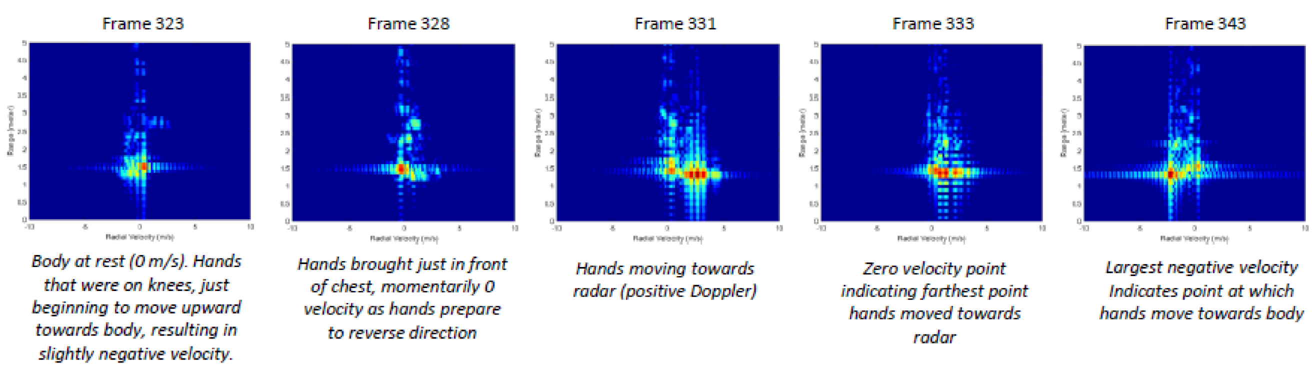

Fig. 3: Samples of sequential frames from Range-Doppler video for the `breathe` sign.

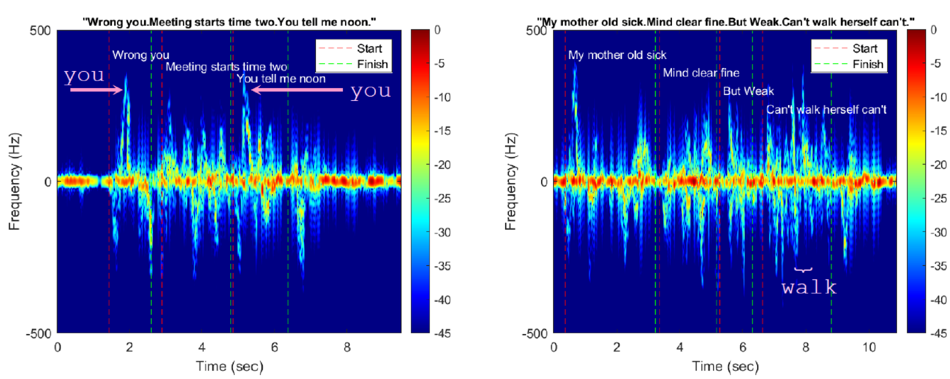

Fig. 4: Samples of the micro-Doppler signature of sequential sentences.

the `breathe` sign, the arms move about 25-30 cm. This is consistent with the radar measurement of 2 range bin displacement. The range resolution of the radar can be improved through transmission of a waveform with greater bandwidth. In this case, the hand displacement could be more accurately measured.

B. Experiment 2: Sequences of Sentences

Although the measurements of individual signs is helpful to gaining a preliminary understanding of the shape and features of motion in ASL, typical signing patterns are greatly affected by *coarticulation*: the effect of previous phenome influencing subsequent phenomes. In sign language, coarticulation occurs because the initial motion of a sign depends or where the hands were located at the end of the previous sign. In our experiments, the participant always began with hands on the knees; however, when the same sign is used mid-sentence, the hands will not be starting from this position. Consider, for example, the sentence `''Wrong you. Meeting starts time two. You tell me noon.''` depicted in Figure 3 includes the sign for `you` twice. This sign is clearly visible in the micro-Doppler signature as the two peaks with the greatest positive velocity or micro-Doppler frequency. The second incidence of `you` is entered rapidly, nearly instantly reaching the peak micro-Doppler frequency, whereas the first incidence has a finite slope reaching the peak frequency. This difference in acceleration is due to ASL coarticulation and is visible in the radar measurements.

As another example, consider the sentence sequence `''My mother sick. Mind clear fine. But weak can't walk herself can't.''` The individual measurement of the sign `walk` reveals that it contains two strong positive and negative peaks together with some smaller frequency components. This signature can be identified in the micro-Doppler signature, but the initial and final frequencies of the sign differ from that seen in the individual measurement, again due to coarticulation and transitions between previous and subsequent signs.

V. Optical Flow from Radar versus Video

Optical flow is a technique often used in computer vision to determine the apparent motion of image objects between two consecutive frames caused by either the movement of the object or camera. When applied to consecutive video frames, the optical flow is a 2D vector field, where each vector is a displacement vector showing the movement of points from one frame to the next. In linguistic studies of ASL, optical flow has been used to extract kinematic metrics. The displacement of an object divided by the time interval between frames thus gives an estimate of object velocity, $\vec{v}_{OFV}$, in the 2D plane perpendicular to the ground.

A. 3D Velocity Estimation via Optical Flow

Optical flow can also be computed from radar data using the range-Doppler videos defined in the previous section. However, in this case, the optical flow carries different physical meaning. Because radar measures range and velocity along the slant range, the optical flow of radar measurements are providing kinematic information along the radial direction (not in azimuth or elevation). Optical flow computed vertically along range indicate the displacement versus time, or radial velocity, $\vec{v}_{OFR}$, whereas computation along the horizontal Doppler axis maps to change in radial velocity versus time, or radial *acceleration*, $\vec{a}_{OFR}$.

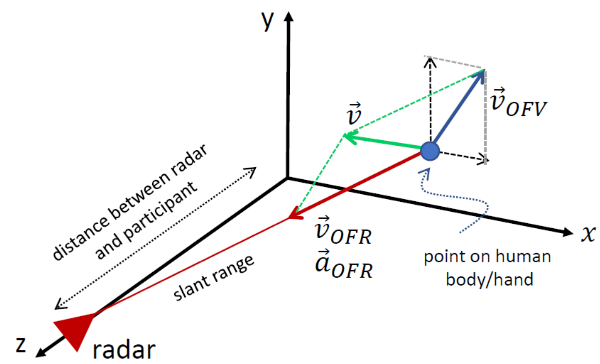

Fig. 5: Geometry of video and radar velocity measurements.

The difference in the geometry of these measurements is shown in Figure 5, while resulting optical flow diagrams for radar and video are compared in Figure 6 for the sign `breathe`. It may be observed that while the velocity estimates from optical flow of video and radar share a similar shape, the radar data exhibit a greater change in pixels per frame than video as well as more texture and detail. Moreover, it is important to note that the velocity estimates of radar and video can be combined to estimate a 3D directional velocity vector, $\vec{v}$. This is a good example of how radar data can be used in fusion with video data to compensate for weaknesses inherent to each modality: radar is effective at measuring radial kinematics, which is difficult for video; however, spatial motion perpendicular to the radar line-of-sight that would difficult to measure with radar is quite easily captured by video. Thus, RF sensors can improve the characterization of ASL kinematics in three-dimensions.

B. Fractal Complexity of Optical Flow

Another way of evaluating the information transfer from frame to frame due to human motion is the *fractal complexity* of the optical flow. First, the power spectral density (PSD) for each velocity is obtained by computing the Fourier transform of the velocity spectrum versus time. This results in a matrix $M(j, f)$ where M is the magnitude, j is a velocity bin, and f is a frequency bin. The fractal complexity, $\beta(j)$, is related to the magnitude at each velocity bin j as follows:

$$M(j, f) = \frac{a}{|f|^{\beta}} \tag{1}$$

where a is an amplitude fitting variable [10]. Taking the logarithm of both sides, a linear expression may be found as

$$ln(M) = ln(a) - \beta ln|f| \tag{2}$$

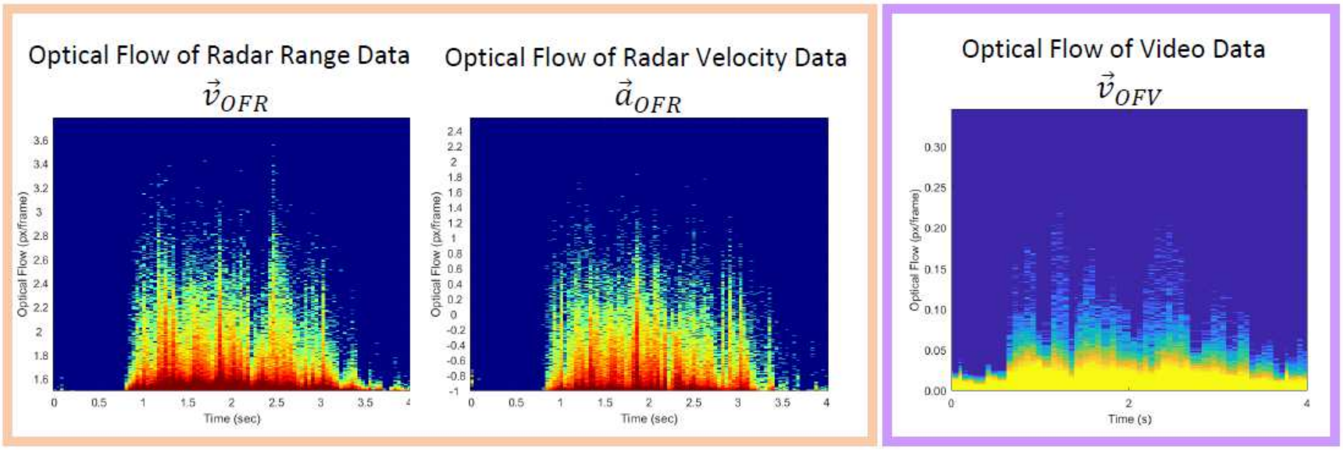

Fig. 6: Optical flow computed from radar and video data for the sign `breathe`.

A simple linear fit is then performed on $ln(f)$ versus $ln(M)$, where β is the slope and a is the intercept on a log–log plot. If $M(j,f)$ is integrated over j, an overall velocity spectrum can be obtained, which, after fitting, then results in an overall fractal complexity, $\bar{\beta}$.

RF measurements from two participants enacting eight different daily activities (building legos, playing chess, cooking, painting, eating, vacuuming, folding laundry, and ironing clothes) were acquired for a duration of 10 minutes. This data was then cropped into 20 second, non-overlapping segments, to yield 30 samples per activity. ASL sentence sequences acquired under Experiment 2 from three different Deaf participants, as tabulated in Figure 2, yielded 3 samples per sentence with approximately 20 second duration. The average $\bar{\beta}$ value of ASL was found as 201.8, while that for daily activities was lower with an average value of 186.8. This reflects the increased information present in the ASL signs due to the presence of communications and underscores the importance of not merely equating ASL with gesturing. Thus, fractal complexity may be a metric useful for identifying periods of communicative signing from motions encountered during the normal course of daily living.

VI. Conclusion

This paper presents initial work exploring different ways in which RF sensors can contribute to linguistic studies of ASL through improved measurements of hand, arm, and upper body kinematics. Our results show RF sensors can also extract estimates of fractal complexity, a metric that has shown a difference in bias between in daily motions and ASL signing periods. In future work, we plan to expand analysis to different aspect angles, different RF frequencies, and exploit fractal complexity for the segmentation of extended duration observations in to intervals of "signing" and "daily activity." This represents an important first step in application of RF sensors for sign language driven smart environments.

VII. Acknowledgements

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

References

[1] M. Erard, "Why sign language gloves don't help deaf people," https://www.theatlantic.com/technology/archive/2017/11/why-sign-language-gloves-dont-help-deaf-people/545441/ , November 2017.

[2] "Information transfer capacity of articulators in american sign language," *Language and speech*, vol. 61, no. 1, pp. 97–112, 2018.

[3] V. Chen, *The Micro-Doppler Effect in Radar*. Artech House, 2010.

[4] C. E. Stilp and K. R. Kluender, "Stimulus statistics change sounds from near-indiscriminable to hyperdiscriminable," *PLOS ONE*, vol. 11, no. 8, 2016.

[5] W. C. Stokoe, "Studies in linguistics: Occasional papers 8," *Sign Language Structure: An Outline of the Visual Communication System of the American Deaf*, 1960.

[6] E. S. Klima, O. J. L. Tzeng, Y. A. Fok, U. Bellugi, D. Corina, and J. G. Bettger, "From sign to script: Effects of linguistic experience on perceptual categorization," *Journal of Chinese Linguistics Monograph Series*, no. 13, pp. 96–129, 1999.

[7] E. Malaia and R. B. Wilbur, "Kinematic signatures of telic and atelic events in asl predicates," *Lang Speech*, vol. 55, no. 3, pp. 407–421, 2012.

[8] E. Malaia, "It still isn't over: Event boundaries in language and perception," *Languistics Compass*, vol. 8, no. 3, pp. 89–98, 2014.

[9] E. Malaia, J. D. Borneman, and R. B. Wilbur, "Assessment of information content in visual signal: analysis of optical flow fractal complexity," *Visual Cognition*, vol. 24, no. 3, p. 246–251, 2016.

[10] J. D. Borneman, E. A. Malaia, and R. B. Wilbur, "Motion characterization using optical flow and fractal complexity," *Journal of Electronic Imaging*, vol. 27, no. 05, July 2018.

[11] E. A. Malaia and R. B. Wilbur, "Syllable as a unit of information transfer in linguistic communication: The entropy syllable parsing model," *Wiley Interdisciplinary Reviews: Cognitive Science*, 2019.

[12] R. G. Bosworth, C. E. Wright, and K. R. Dobkins, "Analysis of the visual spatiotemporal properties of american sign language," *Vision Research*, vol. 164, p. 34–43, Nov. 2019.

[13] B. Erol and M. G. Amin, "Radar data cube processing for human activity recognition using multi subspace learning," *IEEE Transactions on Aerospace and Electronic Systems*, pp. 1–1, 2019.

[14] Y. He, F. Le Chevalier, and A. G. Yarovoy, "Range-doppler processing for indoor human tracking by multistatic ultra-wideband radar," in *2012 13th International Radar Symposium*, May 2012, pp. 250–253.