World Academy of Science, Engineering and Technology
International Journal of Electronics and Communication Engineering
Vol:15, No:3, 2021

Facial Expression Phoenix (FePh): An Annotated Sequenced Dataset for Facial and Emotion-Specified Expressions in Sign Language

Marie Alaghband, Niloofar Yousefi, Ivan Garibay

Abstract—Facial expressions are important parts of both gesture and sign language recognition systems. Despite the recent advances in both fields, annotated facial expression datasets in the context of sign language are still scarce resources. In this manuscript, we introduce an annotated sequenced facial expression dataset in the context of sign language, comprising over 3000 facial images extracted from the daily news and weather forecast of the public tv-station PHOENIX. Unlike the majority of currently existing facial expression datasets, FePh provides sequenced semi-blurry facial images with different head poses, orientations, and movements. In addition, in the majority of images, identities are mouthing the words, which makes the data more challenging. To annotate this dataset we consider primary, secondary, and tertiary dyads of seven basic emotions of "sad", "surprise", "fear", "angry", "neutral", "disgust", and "happy". We also considered the "None" class if the image's facial expression could not be described by any of the aforementioned emotions. Although we provide FePh as a facial expression dataset of signers in sign language, it has a wider application in gesture recognition and Human Computer Interaction (HCI) systems.

Keywords—Annotated Facial Expression Dataset, Sign Language Recognition, Gesture Recognition, Sequenced Facial Expression Dataset.

I. INTRODUCTION

ABOUT 466 million people worldwide have disabled hearing and by 2050 this number will increase to 900 million people–one in every ten (based on the most recent report of World health Organization [1] (WHO) on January 2020). People with hearing loss often communicate through lip-reading skills, texts, and sign language. Sign language, the natural language of people with severe or profound hearing loss, is unique to each country or region [2] and has its own grammar and structure. Combinations of hand movements, postures, and shapes with facial expressions make sign language very unique and complex [3].

Hands are widely used in sign language to convey meaning. Despite the importance of hands in sign language, the directions of eye gazes, eyebrows, eye blinks, and mouths, as part of facial expressions also play an integral role in conveying both emotion and grammar [3]. Facial expressions that support grammatical constructions in sign language help to eliminate the ambiguity of signs [4]. Therefore, sign language recognition systems without facial expression recognition are incomplete [5].

M. Alaghband, N. Yousefi, and I. Garibay are with the Department of Industrial Engineering and Management Systems, University of Central Florida, Orlando, Florida, USA (E-mail: marie.alaghband@knights.ucf.edu, niloofar.yousefi@ucf.edu, and igaribay@ucf.edu).

Due to the importance of hand shapes and movements in sign language, hands are widely used in sign language recognition systems. However, the other integral part of sign language, facial expressions, has not yet been well studied. One reason is the lack of an annotated facial expression dataset in the context of sign language. To the best of our knowledge, an annotated vision-based facial expression dataset in the field of sign language is a very rare resource. This limits researchers' ability to study multi-modal sign language recognition models that consider both facial expressions and hand gestures.

Conducted research in sign language recognition systems can be categorized in two main groups: vision-based and hardware-based recognition systems. Hardware-based recognition systems use datasets that are collected utilizing special colored gloves [8]–[10], special sensors, and/or depth cameras (such as Microsoft Kinect and Leap Motion) [4], [11]–[15], [17], [41] to capture special features of the signer's gestures. Some well-known hardware-based datasets are listed in Table I.

Although utilizing hardware eases the process of capturing special features, they limit the applicability where such hardwares are not available. Therefore, vision-based sign language recognition systems utilizing datasets collected by regular cameras are proposed [2], [6], [18]–[20]. Table II provides a list of some well-known vision-based sign language datasets.

One well-known continuous sign language dataset is the RWTH-PHOENIX-Weather corpus [18]. RWTH-PHOENIX-Weather is a large vocabulary (more than 1200 signs) image corpus containing weather forecasts recorded from German news. Two years later, its publicly available extension called RWTH-PHOENIX-Weather multisigner 2014 dataset was introduced. We will create our annotated facial expression dataset based upon RWTH-PHOENIX-Weather multisigner 2014 continuous sign language benchmark dataset.

Facial expression recognition is a very well established field of research with publicly available databases containing basic universal expressions. CK$^+$ [35] is a well-known facial expression database with 327 annotated video sequences on seven basic universal facial expressions ("anger", "contempt", "disgust", "fear", "happiness", "sadness", and "surprise"). Some other widely used facial expression databases are MMI [36], [37], Oulu-CASIA [38], and FER2013 [39]. FER2013 is an unconstrained large database considering seven emotions (previous six emotions plus "neutral"). Despite the value of these facial expression databases, hardly any one them are in

World Academy of Science, Engineering and Technology
International Journal of Electronics and Communication Engineering
Vol:15, No:3, 2021

TABLE I
A LIST OF SOME HARDWARE-BASED SIGN LANGUAGE DATASETS

Authors	Name	Language	Gesture Type	Language level	Classes	Data Type
[21]	UCI Australian Auslan Sign Language dataset	Australian	Dynamic	Alphabets	95	Data Glove
[11]	ASL Finger Spelling A	American	Static	Alphabets	24	Depth Images
[11]	ASL Finger Spelling B	American	Static	Alphabets	24	Depth Image
[22]	MSRGesture 3D	American	-	Words	12	Depth Video
[23]	CLAP14	Italian	-	Words	20	Depth Video
[24]	ChaLearn LAP IsoGD & ConGD	-	Static & Dynamic	-	249	RGB & Depth Video
[25]	PSL Kinect 30	Polish	Dynamic	Words	30	Kinect Video
[13]	ISL	Indian	static	Alphabets, Numbers, Words	140	Depth Images

TABLE II
AN OVERVIEW OF SOME VISION-BASED SIGN LANGUAGE DATASETS

Authors	Name	Language	Gesture Type	Language level	Classes	Data Type
[28]	-	American	Static	Alphabets, numbers	36	Image
[11]	ASL Finger Spelling A	American	Static	Alphabets	24	Image
[29]	HUST-ASL	American	Static	Alphabets, Numbers	34	RGB & Kinect Image
[19], [26]	Purdue RVL-SLLL ASL Database	American	-	Alphabets, Numbers, Words, Paragraphs	104	Image, Video
[30]	Boston ASLLVD	American	Dynamic	Words	>3300	Video
[20]	ASL-LEX	American	-	Words	Nearly 1000	Video
[2]	MS-ASL	American	Dynamic	-	1000	Video
[31]	-	Arabic	-	Words	23	Video
[18]	RWTH-PHOENIX-Weather 2012	German	-	Sentence	1200	Image
[6], [32]	RWTH-PHOENIX-Weather Multisigner 2014	German	Dynamic	Sentence	1080	Video
[16]	RWTH-PHOENIX-Weather 2014 T	German	Dynamic	Sentence	1066	Video
[27]	SIGNUM	German	-	words, Sentences	450 Words, 780 Sentence	Video
[9]	LSA16	Argentinian	-	Alphabets, Words	16	Image
[9]	LSA64	Argentinian	-	Words	64	Video
[13]	the ISL dataset	Indian	static	Alphabets, Numbers, Words	140	Image
[33]	ISL hand shape dataset	Irish	Static & Dynamic	-	23 Static & 3 Dynamic	Image Video
[34]	Japaneese Finger spelling sign language dataset	Japan	-	-	41	Image

the context of sign language.

In this paper, we introduce FePh, an annotated facial expression dataset for the publicly available continuous sign language dataset RWTH-PHOENIX-Weather 2014 [6].

As a matter of continuity image data, FePh is similar to CK+, MMI, and Oulu-CASIA. However, it is more complex than those databases as it contains real-life captured videos with more than one basic facial expression in each sequence, with different head poses, orientations, and movements.

In addition, FePh not only contains seven basic facial expressions of the FER2013 database, but it also considers their primary, secondary, and tertiary dyads. It is also noteworthy to mention that in sign language, signers mouth the words or sentences to help their audience better grasp meanings. This is a characteristic that makes facial expression datasets in the context of sign language more challenging. As such, this manuscript provides the following contributions: first, introducing annotated facial expression dataset of the RWTH-PHOENIX-Weather 2014 dataset. Second, attributing highly used hand shapes with their associated performed facial expressions, and third, illustrating the relationships between hand shapes and facial expressions in sign language.

II. METHODS

Due to the integral role of facial expressions in conveying emotions and grammar in sign language, it is important to use multi-modal sign language recognition models that consider both hand shapes and facial expressions. Therefore, to create FePh and annotate facial expressions of a sign language dataset with annotated hand shapes, we considered the well-known publicly available continuous RWTH-PHOENIX-Weather 2014 dataset. Since the annotated hand shapes dataset RWTH-PHOENIX-Weather 2014 is publicly available as RWTH-PHOENIX-Weather 2014 MS Handshapes dataset [7], we provide facial expression annotations for the same dataset, which enables researchers to utilize a dataset that has both hand shape and facial expression annotations.

As a starting point, we collected the full frame images of RWTH-PHOENIX-Weather 2014 development set that are identical to the RWTH-PHOENIX-Weather 2014 MS Handshapes dataset [7]. Furthermore, in order to create a solid facial expression dataset and avoid the influence of hand shapes on the facial expression annotators, faces of all full frame images are automatically detected, tracked, and cropped using facial recognition techniques.

Twelve annotators (six women and six men) between 20 to 40 years old were asked to annotate the data. We asked annotators to answer three questions about each static image: the signer's emotion, visibility, and gender. In terms of emotion, annotators could choose one or more of the following applicable basic universal facial expressions for each static image: "sad", "surprise", "fear", "angry", "neutral", "disgust", and "happy". Although more than seven emotions and their primary, secondary, and tertiary dyads exist, considering all of them was not within the scale of this project. Therefore,

World Academy of Science, Engineering and Technology
International Journal of Electronics and Communication Engineering
Vol:15, No:3, 2021

Fig. 1 Exemplary images of obscured faces

TABLE III
THE FACIAL EXPRESSION LABELS WITH THEIR CORRESPONDING NUMBERS

label#	Emotion	label#	Emotion	label#	Emotion
0	neutral	10	anger_neutral	52	sad_disgust
1	anger	21	disgust_anger	53	sad_fear
2	disgust	31	anger_fear	60	neutral_surprise
3	fear	32	fear_disgust	61	surprise_anger
4	happy	40	neutral_happy	62	surprise_disgust
5	sad	50	neutral_sad	63	surprise_fear
6	surprise	51	sad_anger	65	surprise_sad
7	none				

we offered the eighth class of "None" as well. Annotators were asked to choose the "None" class when none of the aforementioned emotions could describe the facial expression of the image. In addition, since annotators could choose more than one facial expression for each individual image, the combinations of basic universal facial expressions were also considered (interestingly, this did not result in choosing more than two emotions for each image) and shown by a "_" in between such as surprise_fear. The sequence of emotions is not important in the secondary and tertiary dyads (i.e., surprise_fear and fear_surprise are the same).

With regard to the second question, visibility, we asked the annotators to evaluate whether the signer's face is completely visible. Although the signer's face was visible in majority of images, this was not always the case. The partial visibility of the face was due to the signer's head movement, position, hand movement, and transitions from one emotion to another emotion. This helped us to detect and opt out these obscured images in the data. Figure 1 shows some obscured exemplary images.

The last question of signer's gender was asked to provide statistics of signers' gender. This statistics enables conducting future researches in the affects of gender in expressing emotions and facial expressions.

For our labelling purpose, we took advantage of the Labelbox [40] annotating solution tool through which we defined an annotation project and randomly distributed images to be labeled by the annotators. In addition, due to the complexities of the facial images of the RWTH-PHOENIX-Weather 2014 dataset, we used the auto consensus option of the Labelbox tool. These complexities are as follows:

- The ambiguity of images, due to signer's movement, head position, and transitions from one emotion to another (e.g., eyes are closed and/or the lips are still open).
- Low quality (resolution) and blurriness of images.
- Mouthed words that confuse facial expression annotators.
- Personal differences between signers expressing facial expressions.
- The best facial expression that describes the image is not included in the dataset.
- Images may not be in facial expression's top frame.
- Large intra-class variance (such as "surprised" emotion with open or closed mouth).
- Inter-class similarities.

With the usage of auto consensus option of Labelbox, we asked more than one annotator (three) to annotate about 60% percent of the data. For the images with three labels, we chose the most voted emotion as the final label of the facial image. In cases where there was not a most voted emotion, but the

image was a part of a sequence of images, we have assigned labels based on the before or after images' facial expression of the same sequence. On the other hand, if there was not a most voted emotion, and the image was not a part of a sequence of images (i.e., one single image without any sequence), we asked our annotators to relabel the image. In this case, all images needed to be labelled by three different annotators.

III. DATA RECORDS

The FePh facial expression dataset produced with the above method, is stored on Harvard Dataverse (https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi: 10.7910/DVN/358QMQ). All facial images are stored in "FePh_images.zip". Although the full frame images of the FePh dataset are identical to the RWTH-PHOENIX-Weather 2014 images, images' filename in the FePh_images.zip are different from the original images. FePh filenames consider both image folder_name and image_file_number. For example, the full frame image of the facial image with filename "01August_2011_Monday_heute_default-6.avi_pid0_fn000054-0.png" in the FePh dataset is identical to the full frame image with the directory of "... / 01August_2011 _Monday_heute_default-6 / 1 / 01August_2011 _Monday_heute.avi_pid0_ fn000054-0.png" in the RWTH-PHOENIX-Weather 2014 and "... / 01August_2011 _Monday_heute_default-6 / 1 /_.png_ fn000054-0.png" image in the RWTH-PHOENIX-Weather 2014 MS Handshapes dataset. This helped us to store all images in one single folder (FePh_images). The FePh_labels.csv file contains images' filenames, facial expression labels, and gender labels. To ease data usability, we stored the facial expression labels as codes. Table III shows the facial expression labels with their corresponding code numbers. In addition, 0 and 1 in the gender column represent the male and female genders, respectively.

IV. TECHNICAL VALIDATION

The FePh dataset is created by manually labelling 3359 images of the RWTH-PHOENIX-Weather 2014 development set that are identical to the full frame images of the RWTH-PHOENIX-Weather 2014 MS Handshapes dataset. Seven universal basic emotions of "sad", "surprise", "fear", "angry", "neutral", "disgust", and "happy" are considered as facial expression labels. In addition to these basic emotions, we asked annotators to choose all the emotions that may apply to an image. This resulted in secondary and tertiary dyads

World Academy of Science, Engineering and Technology
International Journal of Electronics and Communication Engineering
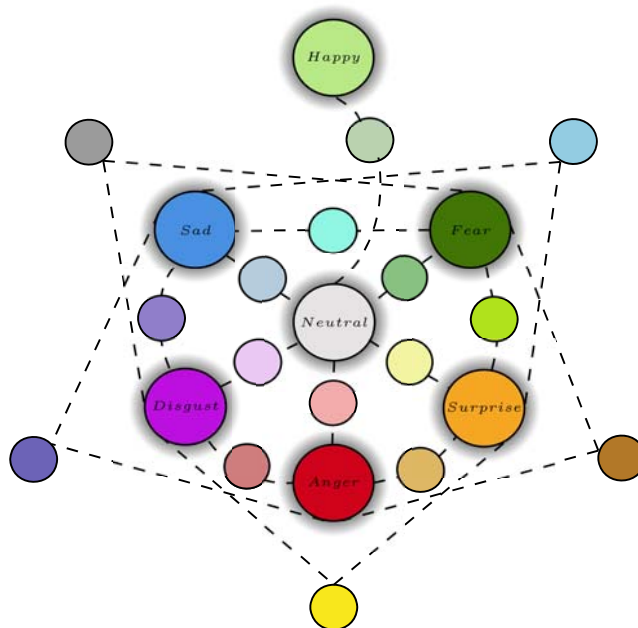Vol:15, No:3, 2021

Fig. 2 Graph displaying the primary, secondary, and tertiary dyads on seven basic universal emotions in the FePh dataset. Seven basic universal emotions are shown with colored circles containing their name. In addition, colored circles that are connecting each two basic emotions show the secondary or tertiary dyads of the basic emotions that they are connected to. For example, the orange circle connected to the "Disgust" and "Surprise" emotions, shows the emotion of "Disgust_Surprise". Emotion "Happy"'s only dyad is with "Neutral" emotion named as "Neutral_Happy".
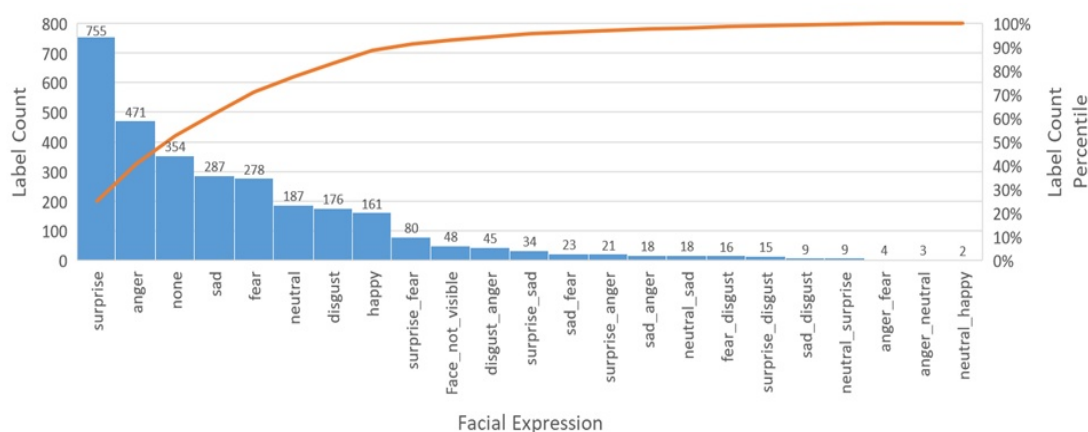
of seven basic emotions such as fear_sad, fear_anger, etc. Interestingly, this did not result in having combinations of three basic emotions. Figure 2 shows the corresponding graph of seven basic emotions, their primary, secondary, and tertiary dydes presented in FePh dataset. Seven basic emotions are shown by colored circles with the emotion labels written inside them. Other colored circles connecting each two basic emotions illustrate the secondary or tertiary dyads of the basic emotions that are connected to. Emotion "Happy" has only one dyad, which is with "Neutral" emotion, named as "Neutral_Happy".

Although the FePh dataset presents annotated facial expression for all hand shape classes of the RWTH-PHOENIX-Weather 2014, we analyzed the facial expression labels for the top 14 hand shapes (i.e., classes "1", "index", "5", "f", "2", "ital", "b", "3", "b_thumb", "s", "pincet", "a", "h", and "ae"). This is due to the demonstrated distribution of the counts per hand shape classes in [7] that shows the top 14 hand shape classes represent 90% of the data. Seven universal basic facial expressions and their secondary or tertiary dyads occur with different frequencies in the data. Figure 3a shows the distribution counts per facial expression class in the data. As it shows, about 90% of the data is expressed with basic facial expressions. In addition, Figures 3b and 3c illustrate the frequency of images with obscured faces and the "None" class in the top 14 hand shape classes, respectively.
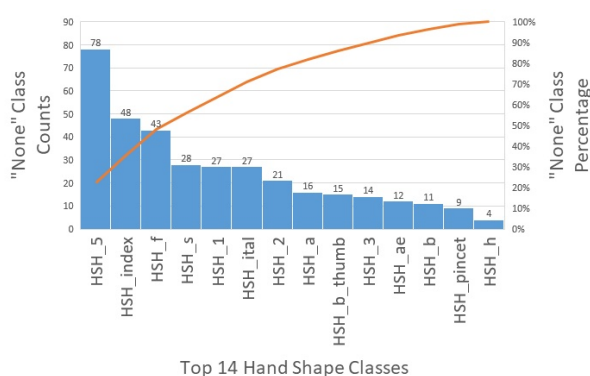
By analyzing facial expressions per hand shape class, we found out that more than one facial expression class represents each hand shape class. Figure 4 shows the frequency heatmaps of the seven facial expressions and their primary, secondary, and tertiary dyads for the top 14 hand shape classes. Each heatmap illustrates the frequency of facial expressions based on the

facial expression graph of seven basic universal emotions and their primary, secondary, and tertiary dyads (shown in Figure 2) for one of the top 14 hand shape classes. The heatmaps show that more than one facial expression is expressed within a single hand shape class, which is due to the complexity of sign language in using facial expressions with hand shapes. Two of these complexities that affect performing different facial expressions within each hand shape class are as follows:
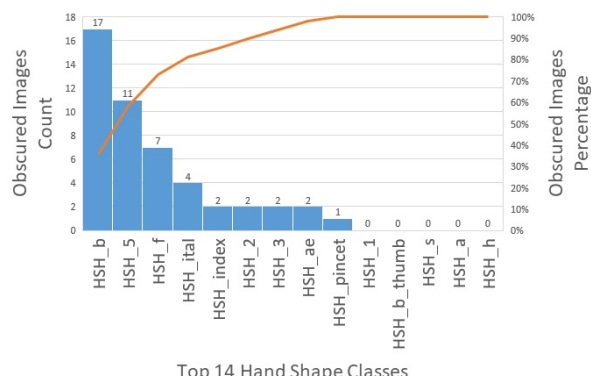
First, although some meanings are communicated using only one hand (usually the right hand), many sign language meanings are communicated using both hands with different hand poses, orientations, and movements. Figure 5a shows some exemplary full frame images of hand shape class "1" of the RWTH-PHOENIX-Weather 2014 dataset with different facial expressions. As the figure illustrates, the usage of right hand shapes have large intra-class variance (i.e., the left hand may not be used or may perform similar or different hand shape from the right hand) that may affect the meanings, and as a result, the facial expressions corresponding to them. The first top row in the figure shows the full frame images with the right hand shape of class "1" and no left hand shape. The second row shows some exemplary full frame images, in which the signer has used both hands. In this row, although both right and left hands demonstrate the same hand shape class (hand shape class "1"), their pose, orientation, and movement can differ, which may affect the corresponding facial emotion. The third row of images in Figure 5a illustrates full frame examples of using both hands with different hand shapes and facial expressions. Therefore, although RWTH-PHOENIX-Weather 2014 MS Handshapes is a valuable resource presenting right hand shape labels, it lacks pose, orientation, and movement labels of the right hand along with the left hand shape labels.

World Academy of Science, Engineering and Technology
International Journal of Electronics and Communication Engineering
Vol:15, No:3, 2021

(a) Pareto chart showing the distribution counts per facial expression class in the FePh dataset for the top 14 hand shape classes (briefly shown as HSH). As the chart shows, basic facial expressions represent 90% of the data.

(b) Pareto chart showing the distribution of the "None" facial expression class counts per top 14 hand shape classes (HSH)

(c) Pareto chart showing the distribution of the obscured image counts per top 14 hand shape classes (HSH)

Fig. 3 Pareto chart showing the distribution of the facial expression "None" and obscured image counts for top 14 hand shape classes (HSH).

Adding this information to the data affects the communicated meanings as well as the facial expressions that are expressed.

Second, due to the communication of grammar via facial expressions, identical hand shapes may be performed with different facial expressions. Figure 5b demonstrates some images of this kind that despite the similarity of hand shapes, the facial expressions are different. This complex usage of hands with large intra-class variance and inter-class similarities help signers to communicate different meanings with similar or different facial expressions.

In addition to the above, the frequency of facial expressions expressed in each hand shape class shows evidence of a meaningful association between hand shapes and facial expressions in the data. To better illustrate this correlation, we calculated the correlation matrix of facial expressions' frequency in each hand shape class. Since the correlations between the facial expressions together is not the focus of this manuscript, the first column of each correlation matrix that shows the correlation between frequency and each facial expression is only considered. Table IV illustrates the first columns of facial expressions and their frequencies of occurrence in the top 14 hand shape classes correlation matrices. Monitoring each column in Table IV gives the most correlated facial expressions for each hand shape. For

example, in the column of hand shape class "3", the positive values of 0.697, 0.383, 0.287, 0.214 and 0.093 (that are in intersections with "fear", "surprise", "anger", "surprise_fear" and "None", respectively) show the positive correlation values with facial expressions in hand shape class "3". These highly correlated facial expressions in each hand shape class can also be interpreted from heatmaps illustrated in Figure 4.

V. Usage Notes

To the best of our knowledge, this dataset is the first annotated vision-based publicly available sequenced facial expression dataset in the context of sign language. Although the number of facial images is enough for statistical and some machine learning methods, it may not be sufficient for some of the state-of-the-art learning techniques in the field of computer vision. Therefore, for such studies, we suggest users to create matched samples choosing subjects from the dataset. This work not only provides an annotated facial expression dataset with different head poses, orientations, and movements, but also contributes in availability of a sign language dataset with both hand shape and facial expression labels with attributions in multi-modal future works in the field. In addition, this dataset has a wider application in other research areas such as gesture recognition and Human-Computer Interaction (HCI) systems.

World Academy of Science, Engineering and Technology
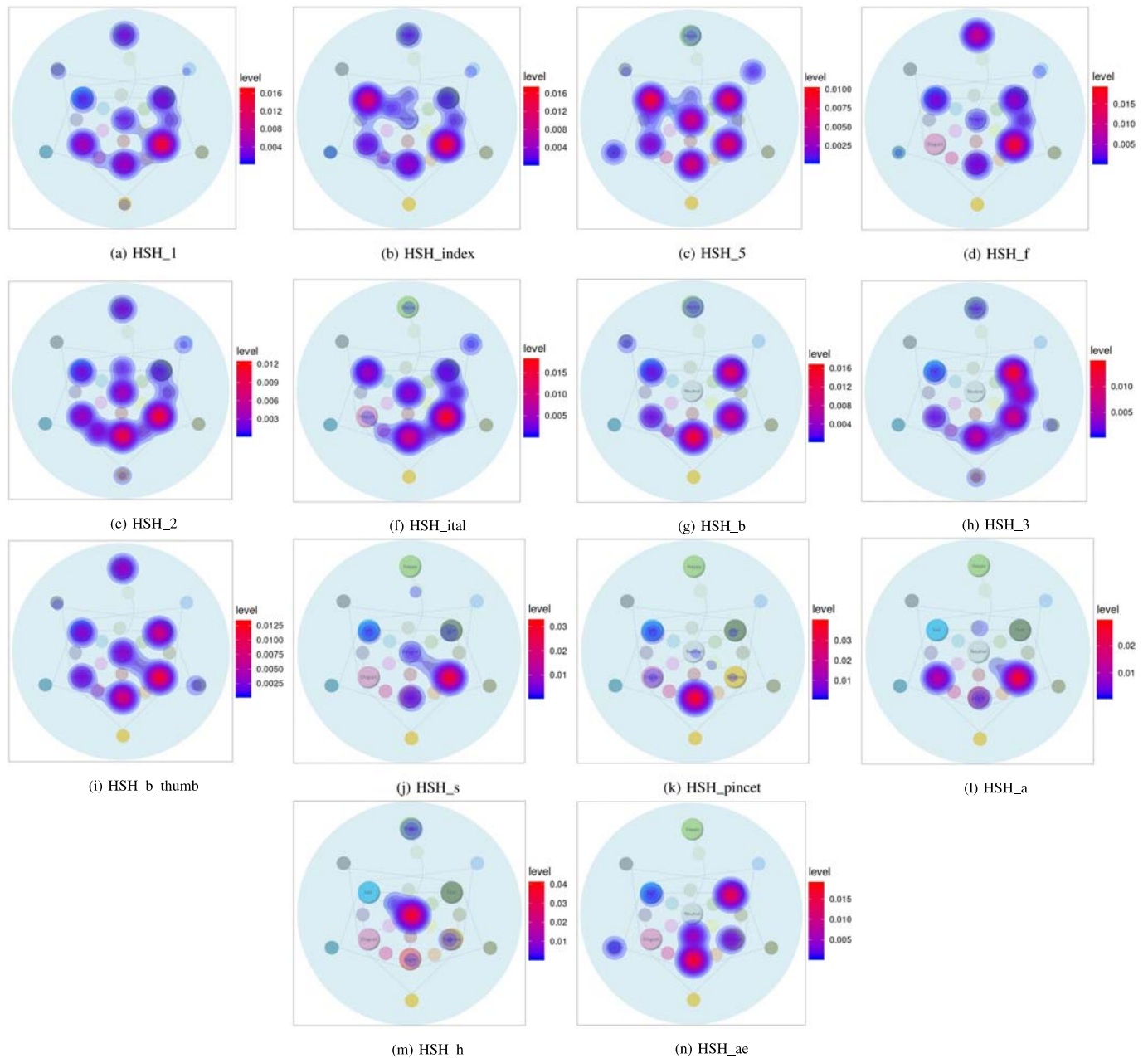International Journal of Electronics and Communication Engineering
Vol:15, No:3, 2021

Fig. 4 Heatmaps showing the frequency distribution of the top 14 hand shape classes. Each heatmap, assigned to one hand shape class (briefly mentioned as HSH), shows the frequency of facial expressions on the assigned hand shape class over the facial expression graph of seven basic universal emotions and their primary, secondary, and tertiary dyads. As they show, more than one facial expressions are expressed within a single hand shape class.

VI. Conclusion

In this work, we presented the FePh dataset, which to the best of our knowledge, is the first real-life annotated sequenced facial expression dataset in the context of sign language. FePh in conjunction with RWTH-PHOENIX-Weather 2014 and RWTH-PHOENIX-Weather 2014 MS Handshapes datasets constitute the first sign language data with both handshapes and facial expression labels. We hope this unique characteristic will propel research in multi-modal sign language and gesture recognition.

As preliminary results and analysis of the data indicate a meaningful relationship between two important modals of sign language (i.e., handshapes and facial expressions), for future work, we propose applying multi-modal learning and computer vision techniques on joint RWTH-PHOENIX-Weather 2014 MS Handshapes and FePh datasets. We believe that the introduction of this dataset will allow the facial expression, sign language, and gesture recognition communities to improve their learning techniques to the latest levels of computer vision trends.

World Academy of Science, Engineering and Technology
International Journal of Electronics and Communication Engineering
Vol:15, No:3, 2021

(a) Exemplary images of hand shape class "1"

(b) Full frame images with similar hand shapes but different facial expressions

Fig. 5 Exemplary full frame images of the RWTH-PHOENIX-Weather 2014 dataset

TABLE IV

CORRELATION MATRIX OF FACIAL EXPRESSIONS' DUMMY VARIABLES AND THEIR FREQUENCIES FOR THE TOP 14 HAND SHAPE CLASSES. EMPTY CELLS SHOW THE ABSENCE OF THE CORRELATION (THE FACIAL EXPRESSION IS NOT EXPRESSED IN THE HAND SHAPE CLASS IMAGES).

| | Frequency of facial expressions for hand shape class | | | | | | | | | | | | | |
	1	index	5	f	2	ital	b	3	b_thumb	s	pincet	a	h	ae
Frequency	1	1	1	1	1	1	1	1	1	1	1	1	1	1
neutral	-0.027	-0.145	0.153	-0.210	0.115	0.066	0.666	-0.221	-0.019	-0.168	-0.150	-0.371	0.995	
anger	0.194	0.114	0.176	-0.047	0.582	0.405	-0.164	0.287	0.343	-0.093	0.991	0.087	-0.259	0.350
disgust	0.233	-0.024	-0.067	0.041	0.075	-0.231	0.459	-0.076	-0.099	-0.242	-0.112			
fear	0.064	-0.089	0.257	0.335	-0.126	-0.061	-0.277	0.697	0.303	-0.192	-0.150			0.250
happy	0.038	-0.080	-0.206	-0.036	-0.045	-0.241	-0.126	-0.124	0.022	0.850	-0.073		-0.209	
sad	-0.040	0.502	0.338	0.805	-0.032	0.130	0.308	-0.124	-0.019	0.379	-0.170			-0.350
surprise	0.857	0.737	0.234	0.139	0.662	0.787	-0.164	0.383	0.624		-0.015	0.612	-0.209	-0.250
none	-0.053	0.187	0.651		-0.005	0.034		0.093	0.263			0.546	-0.109	0.751
anger_neutral														-0.150
disgust_anger	-0.170	-0.153			-0.005	-0.199	-0.353	-0.196	-0.300					
anger_fear								-0.196	-0.260					
fear_disgust	-0.170		-0.230				-0.296		-0.300					
happy_neutral										-0.292				
neutral_sad		-0.097	-0.218										-0.209	
sad_anger		-0.169	-0.160	-0.298				-0.221	-0.300					-0.350
sad_disgust		-0.169	-0.195											
sad_fear		-0.145	-0.206		-0.139							-0.437		
neutral_surprise	-0.222									-0.260	-0.242	-0.150	-0.437	
surprise_anger	-0.209				-0.192	-0.178		-0.148						
surprise_disgust	-0.183				-0.219			-0.172						
surprise_fear	-0.118	-0.113	-0.241	-0.189	-0.192	-0.125		0.214						
surprise_sad	-0.190	-0.169			-0.287	-0.219	-0.178							

REFERENCES

[1] World health organization, deafness and hearing loss, online, 2020. [online]. available: https://www.who.int/healthtopics/hearing-loss.
[2] Joze, H. R. V., & Koller, O. (2018). Ms-asl: A large-scale data set and benchmark for understanding american sign language. arXiv preprint arXiv:1812.01053.
[3] Neiva, D. H., & Zanchettin, C. (2018). Gesture recognition: a review focusing on sign language in a mobile context. Expert Systems with Applications, 103, 159-183.
[4] Freitas, F. A., Peres, S. M., Lima, C. A., & Barbosa, F. V. (2017). Grammatical facial expression recognition in sign language discourse: a study at the syntax level. Information Systems Frontiers, 19(6), 1243-1259.
[5] Kumar, P., Roy, P. P., & Dogra, D. P. (2018). Independent bayesian classifier combination based sign language recognition using facial expression. Information Sciences, 428, 30-48.
[6] Koller, O., Forster, J., & Ney, H. (2015). Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers. Computer Vision and Image Understanding, 141, 108-125.
[7] Koller, O., Ney, H., & Bowden, R. (2016). Deep hand: How to train a cnn on 1 million hand images when your data is continuous and weakly labelled. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3793-3802).
[8] Kadous, M. W. (2002). Temporal classification: Extending the classification paradigm to multivariate time series. Kensington: University of New South Wales.
[9] Ronchetti, F., Quiroga, F., Estrebou, C. A., Lanzarini, L. C., & Rosete, A. (2016). LSA64: an Argentinian sign language dataset. In XXII Congreso Argentino de Ciencias de la Computación (CACIC 2016)..
[10] Wang, R. Y., & Popović, J. (2009). Real-time hand-tracking with a color glove. ACM transactions on graphics (TOG), 28(3), 1-8.
[11] Pugeault, N., & Bowden, R. (2011, November). Spelling it out: Real-time ASL fingerspelling recognition. In 2011 IEEE International conference on computer vision workshops (ICCV workshops) (pp. 1114-1119). IEEE.
[12] Ren, Z., Yuan, J., & Zhang, Z. (2011, November). Robust hand gesture recognition based on finger-earth mover's distance with a commodity depth camera. In Proceedings of the 19th ACM international conference on Multimedia (pp. 1093-1096).
[13] Ansari, Z. A., & Harit, G. (2016). Nearest neighbour classification of Indian sign language gestures using kinect camera. Sadhana, 41(2), 161-182.
[14] Zafrulla, Z., Brashear, H., Starner, T., Hamilton, H., & Presti, P. (2011, November). American sign language recognition with the kinect. In

World Academy of Science, Engineering and Technology
International Journal of Electronics and Communication Engineering
Vol:15, No:3, 2021

Proceedings of the 13th international conference on multimodal interfaces (pp. 279-286).

[15] Uebersax, D., Gall, J., Van den Bergh, M., & Van Gool, L. (2011, November). Real-time sign language letter and word recognition from depth data. In 2011 IEEE international conference on computer vision workshops (ICCV Workshops) (pp. 383-390). IEEE.

[16] Cihan Camgoz, N., Hadfield, S., Koller, O., Ney, H., & Bowden, R. (2018). Neural sign language translation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 7784-7793).

[17] Mehdi, S. A., & Khan, Y. N. (2002, November). Sign language recognition using sensor gloves. In Proceedings of the 9th International Conference on Neural Information Processing, 2002. ICONIP'02. (Vol. 5, pp. 2204-2206). IEEE.

[18] Forster, J., Schmidt, C., Hoyoux, T., Koller, O., Zelle, U., Piater, J. H., & Ney, H. (2012, May). RWTH-PHOENIX-Weather: A Large Vocabulary Sign Language Recognition and Translation Corpus. In LREC (Vol. 9, pp. 3785-3789).

[19] Martínez, A. M., Wilbur, R. B., Shay, R., & Kak, A. C. (2002, October). Purdue RVL-SLLL ASL database for automatic recognition of American Sign Language. In Proceedings. Fourth IEEE International Conference on Multimodal Interfaces (pp. 167-172). IEEE.

[20] Caselli, N. K., Sehyr, Z. S., Cohen-Goldberg, A. M., & Emmorey, K. (2017). ASL-LEX: A lexical database of American Sign Language. Behavior research methods, 49(2), 784-801.

[21] Kadous, W. (1995). Grasp: Recognition of Australian sign language using instrumented gloves.

[22] Kurakin, A., Zhang, Z., & Liu, Z. (2012, August). A real time system for dynamic hand gesture recognition with a depth sensor. In 2012 Proceedings of the 20th European signal processing conference (EUSIPCO) (pp. 1975-1979). IEEE.

[23] Escalera, S., Baró, X., Gonzalez, J., Bautista, M. A., Madadi, M., Reyes, M., ... & Guyon, I. (2014, September). Chalearn looking at people challenge 2014: Dataset and results. In European Conference on Computer Vision (pp. 459-473). Springer, Cham.

[24] Wan, J., Zhao, Y., Zhou, S., Guyon, I., Escalera, S., & Li, S. Z. (2016). Chalearn looking at people rgb-d isolated and continuous datasets for gesture recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 56-64).

[25] Kapuscinski, T., Oszust, M., Wysocki, M., & Warchol, D. (2015). Recognition of hand gestures observed by depth cameras. International Journal of Advanced Robotic Systems, 12(4), 36.

[26] Wilbur, R., & Kak, A. C. (2006). Purdue RVL-SLLL American sign language database.

[27] Von Agris, U., Knorr, M., & Kraiss, K. F. (2008, September). The significance of facial features for automatic sign language recognition. In 2008 8th IEEE International Conference on Automatic Face & Gesture Recognition (pp. 1-6). IEEE.

[28] Barczak, A. L. C., Reyes, N. H., Abastillas, M., Piccio, A., & Susnjak, T. (2011). A new 2D static hand gesture colour image dataset for ASL gestures.

[29] Feng, B., He, F., Wang, X., Wu, Y., Wang, H., Yi, S., & Liu, W. (2016). Depth-projection-map-based bag of contour fragments for robust hand gesture recognition. IEEE Transactions on Human-Machine Systems, 47(4), 511-523.

[30] Athitsos, V., Neidle, C., Sclaroff, S., Nash, J., Stefan, A., Yuan, Q., & Thangali, A. (2008, June). The american sign language lexicon video dataset. In 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (pp. 1-8). IEEE.

[31] Shanableh, T., Assaleh, K., & Al-Rousan, M. (2007). Spatio-temporal feature-extraction techniques for isolated gesture recognition in Arabic sign language. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 37(3), 641-650.

[32] Forster, J., Schmidt, C., Koller, O., Bellgardt, M., & Ney, H. (2014, May). Extensions of the Sign Language Recognition and Translation Corpus RWTH-PHOENIX-Weather. In LREC (pp. 1911-1916).

[33] Oliveira, M., Chatbri, H., Ferstl, Y., Farouk, M., Little, S., O'Connor, N. E., & Sutherland, A. (2017). A dataset for irish sign language recognition.

[34] Hosoe, H., Sako, S., & Kwolek, B. (2017, May). Recognition of JSL finger spelling using convolutional neural networks. In 2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA) (pp. 85-88). IEEE.

[35] Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010, June). The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In 2010 ieee computer society conference on computer vision and pattern recognition-workshops (pp. 94-101). IEEE.

[36] Pantic, M., Valstar, M., Rademaker, R., & Maat, L. (2005, July). Web-based database for facial expression analysis. In 2005 IEEE international conference on multimedia and Expo (pp. 5-pp). IEEE.

[37] Valstar, M., & Pantic, M. (2010, May). Induced disgust, happiness and surprise: an addition to the mmi facial expression database. In Proc. 3rd Intern. Workshop on EMOTION (satellite of LREC): Corpora for Research on Emotion and Affect (p. 65).

[38] Zhao, G., Huang, X., Taini, M., Li, S. Z., & Pietikälnen, M. (2011). Facial expression recognition from near-infrared videos. Image and Vision Computing, 29(9), 607-619.

[39] Goodfellow, I. J., Erhan, D., Carrier, P. L., Courville, A., Mirza, M., Hamner, B., ... & Zhou, Y. (2013, November). Challenges in representation learning: A report on three machine learning contests. In International Conference on Neural Information Processing (pp. 117-124). Springer, Berlin, Heidelberg.

[40] Labelbox, "labelbox," online, 2019. [online]. available: https://labelbox.com.

[41] de Almeida Freitas, F., Peres, S. M., de Moraes Lima, C. A., & Barbosa, F. V. (2014, May). Grammatical facial expressions recognition with machine learning. In The Twenty-Seventh International Flairs Conference.

Marie Alaghband is a qualified PhD student of Industrial Engineering department and a member of the Complex Adaptive Systems Laboratory (CASL) at the University of Central Florida (UCF). Marie received her M.Sc. in Socio-economic systems Engineering (The Institute for Management and Planning Studies; 2013-2016), and a B.Sc. in Statistics (Isfahan University of Technology; 2007-2012). She is currently a graduate teaching and research assistant in the Industrial Engineering and Management Systems (IEMS) department at UCF. Her research expertise lies in scheduling, operation research, and economics, while in her PhD are machine learning and data analysis in Industrial engineering related topics.
    Research Interest: Machine Learning, Computer Vision, Deaf Education and Study, Data Analysis

Niloofar Yousefi is the science director for the UCF Complex Adaptive Systems (CASL) Lab and a post-doctoral research associate in Machine Learning and Computational Learning Theory. She received her Ph.D. in Industrial Engineering/Operation Research with an emphasis in Computer Science, Machine Learning, and Pattern Recognition. She has a Master of Science in Operation Research from University of Tehran and a Bachelor of Science in Applied Mathematics from Iran University of Science and Technology.
    Research Interest: Machine Learning, Statistical Learning Theory, Pattern Recognition, Kernel-based Models, Multi-Task Learning

Ivan Garibay serves as the director of the Complex Adaptive Systems Laboratory (CASL) and the Master of Science in Data Analytics (MSDA) program at the University of Central Florida (UCF). Dr. Garibay received his B.Sc. and Diploma in EE (Ricardo Palma University; 1994,1995), and MS, Ph.D. in CS (University of Central Florida; 2000, 2004). He is currently an Assistant Professor in the Industrial Engineering and Management Systems (IEMS) department at UCF. His research expertise lies in complex systems, agent-based models, data and network science, artificial intelligence and machine learning. His research is currently sponsored by federal agencies and industry by more than 7.0 M. He has published and presented more than 75 papers in journals and conferences.
    Research Interest: Complex systems, Computational Social Science, Artificial Intelligence, Innovation