

# Modeling and Synthesizing Idiopathic Facial Paralysis

Maryam Moosaei, Maryam Pourebadi, and Laurel D. Riek

Computer Science and Engineering, UC San Diego

**Abstract**— Over 22 million people worldwide are affected by Parkinson’s disease, stroke, and Bell’s palsy (BP), which can cause facial paralysis (FP). People with FP have trouble having their expressions understood: both laypersons and clinicians have difficulty understanding them and often misinterpret them, which can result in poor social interactions and poor care delivery. One way to address this problem is through better education and training, of which computational tools may prove invaluable. Thus, in this paper, we explore how to build systems that can recognize and synthesize asymmetrical facial expressions. We introduce a novel computational model of asymmetric facial expressions for BP, which we can synthesize on either virtual and robotic patient simulators. We explore this within the context of clinical education, and built a patient simulator with synthesized FP in order to help clinicians perceive facial paralysis in patients. We conducted both computational and human-focused evaluations of the model, including the feedback from clinical experts. Our results suggest that our BP model is realistic, and comparable to the expressions of people with BP. Thus, this work has the potential to provide a practical training tool for clinical learners to better understand the expressions of people with BP. Our work can also help researchers in the facial recognition community to explore new methods for asymmetric facial expression analysis and synthesis.

## I. INTRODUCTION

Every year, 22 million people experience stroke, Parkinson’s disease, Moebius syndrome, and Bell’s Palsy (BP) [1], [22], [40], which can cause facial paralysis (FP). FP is the inability to move one’s facial muscles on the affected side of the face, leading to asymmetric facial expressions [6]. The quality of social interaction that people with asymmetric facial expressions experience can be poor due to others who have difficulty understanding their emotions [9]. Studies show observers perceive the emotions of a person with FP differently from their actual emotional states [37]. For example, people with severe FP are perceived as less happy than people with mild FP [10]. For people with Parkinson’s disease, observers may mistake expressions of happiness, as signifying depression or deception [2], [38].

In clinical contexts, these misperceptions can lead to poor care delivery. Healthcare providers frequently have negatively biased impressions of patients with facial nerve paralysis [39], which may adversely affect the quality of care they receive [35], [36]. If a patient and a healthcare provider do not communicate effectively, there is a higher chance that their treatment will be unsuccessful [2], [37]. Therefore, new training tools which enable clinical learners to practice their interaction with FP patients may result in improved care for people with FP, and also improve how clinicians calibrate their perception of asymmetric expressions.

Virtual and robotic patient simulators are one of the most commonly used training tools in clinical education [34]. They provide clinical learners with a low-risk, high-fidelity learning environment to practice their procedural and communication skills [29]. Robotic patient simulators (RPS), in particular, can convey realistic, immersive training experiences for learners. They are lifelike, patient-sized humanoid robots that can simulate human physiological responses.

Research suggests that using these simulators may reduce preventable medical errors, which cause approximately 400,000 deaths per year in the US hospitals alone [19], [25]. However, current commercial simulators suffer from a major design flaw: they completely lack facial expressions (see Figure 1). Our team has created expressive virtual and robot patient simulators, which show promise as an important clinical education tool [32], [25], [28], [27], [33]. The development of these simulators was based on the assumption that human faces are structurally symmetric. However, due to the large number of people affected by FP, it is also important to explore the synthesis of asymmetric facial expressions in these contexts. To our knowledge, FP patient simulators have not been explored in this way. Employing simulators in this way may help providers avoid forming biased impressions, improve clinical communication, and, therefore, improve care delivery for people with FP.

In this paper, we introduce the concept of using masked synthesis on patient simulators in order to model asymmetric facial expressions, situated within a clinical education context. Masks are computational models derived from recognized expressions of real people with FP. Masked synthesis is a process of using pre-built masks based on the face of a person with FP, and overlaying it on the stream of standard performance driven synthesis to recreate the asymmetric facial expressions [27]. The longitudinal goal of our research is to build accurate models of people with asymmetric facial expressions, and to help support clinical engagement with people who have FP.

The contributions of this paper are twofold. First, we present a novel algorithm to build accurate computational models (masks) of people with BP that are constructible in real time (See Section III). Second, we applied the algorithm to synthesize BP on virtual patients, and found that clinicians perceive it to be realistic and comparable to humans with BP (See Sections IV and V).

This work is important for the greater affective computing and patient simulation communities because it allows researchers to explore new methods for synthesizing facial expressions. Moreover, by leveraging the BP patient simulator approach presented in this paper, clinical learners may

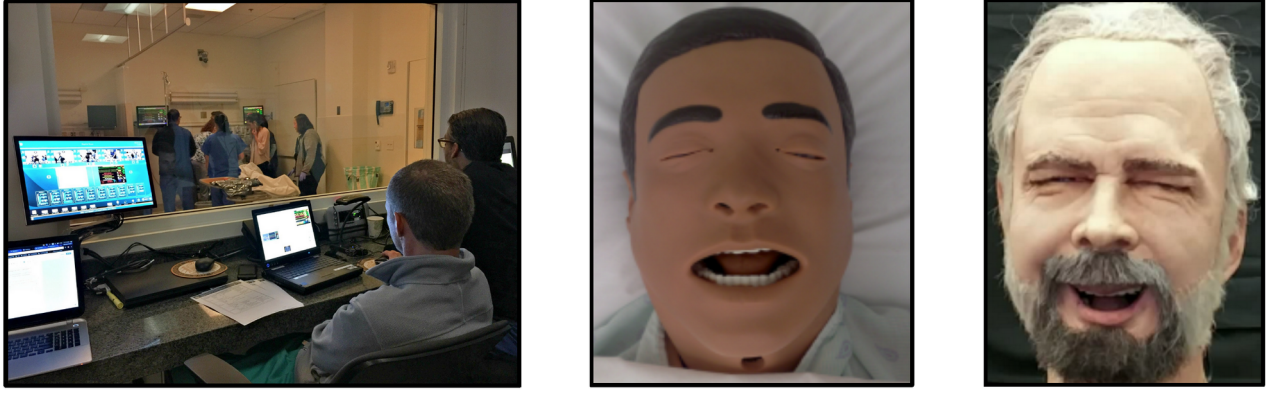


Fig. 1. Left: The simulation center setup where a team of clinical learners treat a non-expressive HPS, which is controlled by simulation operators in a control room. Center: a commonly used inexpressive mannequin head. Right: An example of an expressive RPS system our team built, synthesizing pain.

have the potential to more accurately diagnose people with BP, and to be better able to interact with them. We discuss the implications of these findings in Section VI)

## II. BACKGROUND

Recognizing and synthesizing facial expressions is desirable for a variety of different applications including: human face and head modeling [11], [14], illofacial surgery [8], and rendering robot faces [20]. While there is a significant body of literature exploring symmetric expressions [14], [8], it is also important to study asymmetric facial expressions.

Researchers have had success in identifying the salient features of asymmetric and restricted facial expressions. For example, Tickle-Degnen et al. [42] designed a study to identify reliable emotional cues from expressive behavior in women and men with Parkinson’s disease. Other researchers proposed different quantitative analysis methods to measure the facial asymmetry of facial images [5], [21], [30], [31].

While previous work has laid the foundation for exploring asymmetry, it is critical to study synthesizing asymmetry, especially in clinical education settings. People with asymmetric facial expressions have limited facial expressivity, which makes it difficult for others to form a reliable understanding of their emotions. Moreover, in clinical settings, if a patient and provider cannot communicate effectively, it can adversely affect rapport with the patient and their care decisions [2], [37].

Building models to synthesize asymmetric facial expressions on virtual or physical simulator faces may help to improve the social and procedural skills of clinicians and help promote the quality of care they give to patients with BP. Many researchers have worked on both facially expressive virtual simulators [11] and expressive physical robots [32], [26], [25], [28], [27]. Still, there is a lack of work done on developing patient simulators capable of expressing asymmetric facial expressions.

One of the most commonly used modalities in clinical simulation centers are virtual and robotic patient simulators. Patient simulators help improve clinicians’ procedural and communication skills and enable them to provide effective treatment to patients [28]. These systems provide caregivers

and clinical learners with a low-risk, high-fidelity, clinically-similar learning environment to practice their skills [29]. Although using simulators may reduce preventable medical errors [19], the absence of facial expressions on these simulators may adversely affect patient outcomes [25], [23].

To address this issue, we have built both virtual and robotic patient simulators able to express a range of pathologies, including pain and stroke [32], [25]. We also introduced a generalized automatic framework that can accurately map facial expressions from a performer’s face to both simulated and robotic faces in real-time [27], [28]. The method is based on performance-driven synthesis, which maps motions from video of an operator/educator onto the face of an embodiment (e.g., virtual avatar or robot). In our current work, we build on this to explore a new avenue: the recognition and synthesis of asymmetric facial expressions.

## III. METHODOLOGY

In our work, we are interested in a particular type of FP, Bell’s Palsy (BP). We explore two main research questions in this work. First, how can we computationally model the facial characteristics of BP and synthesize them on a simulator? This is an important question because answering it would enable the development simulators capable of expressing asymmetric facial expressions. Second, how realistically does our mask model convey signs of BP when synthesized on a virtual patient? Addressing this question will help inform the potential clinical efficacy of such an approach using FP simulators as clinical educational tools.

We developed a new masked synthesis method for asymmetric facial expressions, and addressed the aforementioned research questions by engaging in the following activities. First, we acquired videos of people with BP and extracted facial features using a CLM-based approach (See Section III-A). Next, we built computational models (masks) representing the facial characteristics of BP. We then overlaid these prebuilt masks onto a stream of facial expressions generated by standard performance-driven synthesis (See Fig. 2). Next, we transferred the generated asymmetric expressions to the face of a virtual patient simulator (See Section III-B). Finally, we ran an expert-based study to evaluate the realism of the

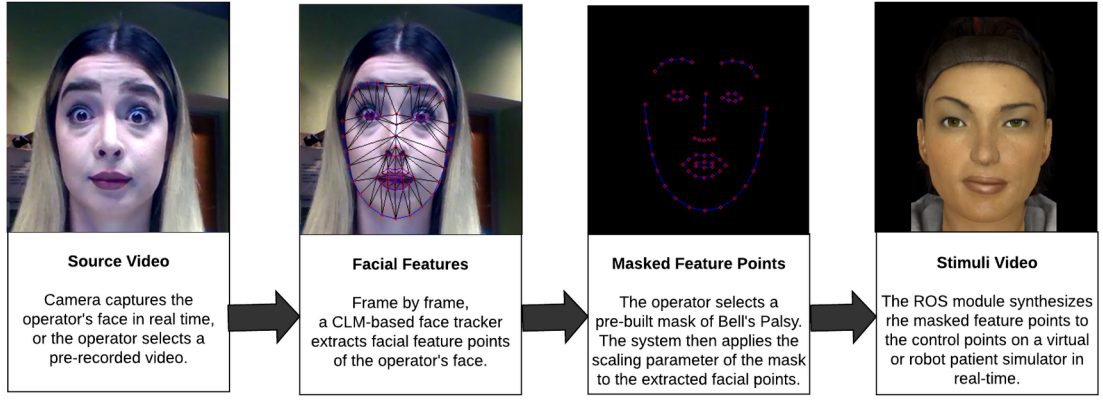


Fig. 2. The context for performing masked synthesis. This can be performed on either a virtual or robotic patient simulator.

synthesized expressions in comparison to actual patients (See Section IV).

#### A. BP video acquisition and facial feature extraction

We focused on one example pathology, BP which affects the facial nerve, causing facial weakness and an inability to control the affected side of the face. The facial weakness usually involves eyes, mouth, and forehead (See Fig. 5).

To build computational models of the facial characteristics of asymmetric expressions, the first step was to acquire source videos from people with BP expressing a wide range of expressions. Therefore, we collected self-recorded, publicly-available videos from people with BP on YouTube.

Many people who have experienced BP have recorded videos of their experience from their diagnosis to their recovery. In these videos, people convey a wide range of expressions including raising their eyebrow, furrowing their brow, smiling, and closing their affected eye to show how BP affects these expressions and how the condition improves over time. Figure 5 presents some example frames from a source video downloaded from YouTube, and shows how BP affects different facial movements.

We downloaded ten source videos from YouTube that present people with BP conveying four expressions (raising eyebrow, furrowing brow, smiling, and closing the eye) required for assessing BP. To ensure that these videos have BP, we only downloaded videos in which either people verbally state their BP diagnosis or the video contains a textual tag indicating BP.

We processed all the source videos to include only ones of high quality in our work (e.g., limited noise and occlusion). Additionally, since the source videos had different lengths, we pre-processed them by cutting them to the same length and only including parts where people were conveying facial expressions.

To extract facial features, we used a CLM-based face tracker implemented by Baltrusitis et al. [3] to track 68 facial points frame-by-frame from each of the source videos (see Fig. 3). We then asked two human judges to watch all the

source videos and label the ones that the face tracker can accurately track. We only included the well-tracked videos in our work.

#### B. Build computational models of Bell's palsy

We built a model for BP inspired by the methods used in the virtual animation literature. Particularly, we have developed our model based on the work by Boker et al. [11], who performed research on manipulating facial expressions and head movements in real time during face to face conversation in video conferencing. Their experiment included two conversants who were sitting in separate rooms, each facing a camera. The researchers used Active Appearance Models (AAMs) [12] to track the facial features of the first conversant. They modified the tracked facial expressions of the first conversant and used them to re-synthesize an avatar with the modified expressions.

The avatar interacted with the second conversant in a video conference. Second conversants were asked if they noticed anything unusual during the conversation. None of the participants noticed they were talking to a computer-generated avatar and none of them guessed that the expressions were manipulated.

To modify the expressions, Boker et al. [11] used an AAM-based face tracker. In Active Appearance Models, the shape of the face is defined by a number of landmarks and their interconnectivity, which shapes a triangulated mesh. In other words, shape  $S$  is defined by 2D coordinates of  $n$  facial landmarks:  $S = [x_1, y_1, x_2, y_2, \dots, x_n, y_n]^T$ . These facial landmarks are vertices of the triangulated mesh.

The training step of AAM models requires manually labeling facial landmarks in some frames and running Principal Component Analysis (PCA) to find the shape model of a face. After the training step, the shape model is defined as below:

$$S = S_0 + \sum_{i=1}^m P_i S_i \quad (1)$$

$S_0$  is the mean shape and the  $S_i$  vectors are the  $m$  vectors that show the shape variation of facial landmarks (mesh

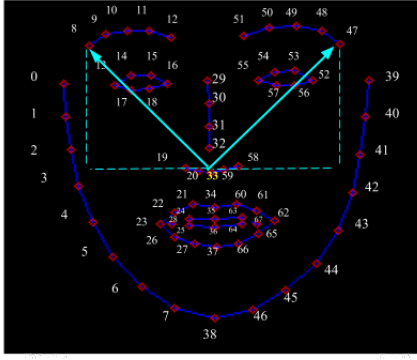


Fig. 3. Comparing the calculated coordinates of feature points with the actual coordinates of the points on the affected side of the face

vertices) from the mean. The coefficients  $P_i$  are the shape parameters which define the contribution of each of the shape vectors  $S_i$  in the shape  $S$ . Therefore, scaling the  $P_i$  parameters will exaggerate or attenuate the facial expressions.

$$S = S_0 + \sum_{i=1}^n P_i S_i \beta_i \quad (2)$$

$\beta$  is a scalar which will exaggerate the expressions if it is bigger than one and will attenuate the expressions if it is smaller than one. Boker et al. [11] used  $\beta$  values smaller than one to reduce expressiveness of a conversant. They wanted to study if changing the expressiveness would affect the behavior during a conversation.

We used a similar approach to build a model for BP. However, in our work, we used a CLM-based face tracker to track 68 facial points. Constrained Local Models (CLM) [13] are similar to Appearance Local Models (AAM) [12] except they do not require manual labeling, and therefore are an improvement over AAMs. Similar to Boker et al. [11], we scaled the CLM parameters by using equation 2. However, instead of attenuating the parameters like Boker et al. [11], we scaled them based on the scaling parameters that we calculated for a pathology (BP) in the following section.

The CLM-based face tracker that we used in our work tracks 68 facial points frame-by-frame (See Fig. 3). These 68 facial points are vertices of a triangulated mesh. To build the model of a pathology (e.g. BP), we found the 68 scaling parameters for the facial feature points for that pathology. We did this in three steps which are explained below.

First, we selected a well-tracked source video that included a wide range of facial expressions expressed by a person with BP. From the video, we selected 20 frames in which the asymmetry is most noticeable.

Second, using the CLM-based face tracker, we tracked 68 facial points in each of the 20 chosen frames. We preprocess the feature points to remove the effect of translation, rotation, and distance to the camera. Then, using the 2D coordinates of the 34 facial features of the unaffected side of the face, we calculated the 2D coordinates of the other part of the face, assuming that the person did not have asymmetric facial expressions. To do so, we used the tip of the nose

as a reference point. The tip of the nose stays static through transitioning from one facial expression to the other.

In a person without FP, each facial feature of one side of the face has the same distance to the tip of the nose as its corresponding point on the other side of the face. For example, the left eye corner and the right eye corner each have the same distance to the tip of the nose. In a person with asymmetric facial expressions, this is not true. Therefore, we used the distance from the tip of the nose to calculate the 2D coordinates of 34 feature points on the affected side of the face, assuming the person did not have FP.

Without loss of generality, assume that the left side of the face is affected by FP. In each frame, we track 2D coordinates of 68 facial features. As seen in Fig. 3, there are 29 feature points on each side of the face and 10 feature points on the line symmetry of a face. Feature points (0, ..., 28) are the 29 points on the right side of the face, feature points (29, ..., 38) are the 10 feature points on the line of symmetry of the face, and feature points (39, ..., 67) are the 29 feature points on the left side of the face. We used feature point 33 (the tip of the nose) as a reference point.

After removing the effect of translation and rotation in each frame,  $R_x$  and  $R_y$  in 3 are two arrays of  $x$  and  $y$  coordinates of the 29 feature points on the right (unaffected) side of the face.  $L_x$  and  $L_y$  in 4 are two arrays of  $x$  and  $y$  coordinates of the 29 feature points on the left (affected) side of the face.

$$R_x = [x_0, x_1, \dots, x_{28}], R_y = [y_0, y_1, \dots, y_{28}] \quad (3)$$

$$L_x = [x_{39}, x_{40}, \dots, x_{67}], L_y = [y_{39}, y_{40}, \dots, y_{67}] \quad (4)$$

$L_x^0$  and  $L_y^0$  are the two arrays with the estimated  $x$  and  $y$  coordinates of the left (affected) side of the face assuming the person did not have asymmetric facial expressions.

$$L_x^0 = [x_{38}^0, x_{39}^0, \dots, x_{65}^0], L_y^0 = [y_{38}^0, y_{39}^0, \dots, y_{65}^0] \quad (5)$$

As Fig. 3 presents:

$$y_i^0 = y_{i-38} \quad 37 < i < 66 \quad (6)$$

$$x_i^0 = (x_{32} - x_{i-38}) + x_{32} = 2x_{32} - x_{i-38} \quad 37 < i < 66 \quad (7)$$

Third, we compared the calculated coordinates of points from the second step ( $L_x^0$  and  $L_y^0$ ) with the actual coordinates of the feature points on the affected side of the face ( $L_x$  and  $L_y$ ). Dividing the actual coordinates by the calculated coordinates gave us scaling parameter for  $x$  and  $y$  coordinates of each of the facial points. Therefore, the scaling parameters,  $\beta_{i,x}$  and  $\beta_{i,y}$  for  $x$  and  $y$  coordinates of each facial point on the left side of face will be calculated as below:

$$\beta_{i,x} = \frac{x_i^0}{x_i} \beta_{i,y} = \frac{y_i^0}{y_i}, \quad 37 < i < 66 \quad (8)$$



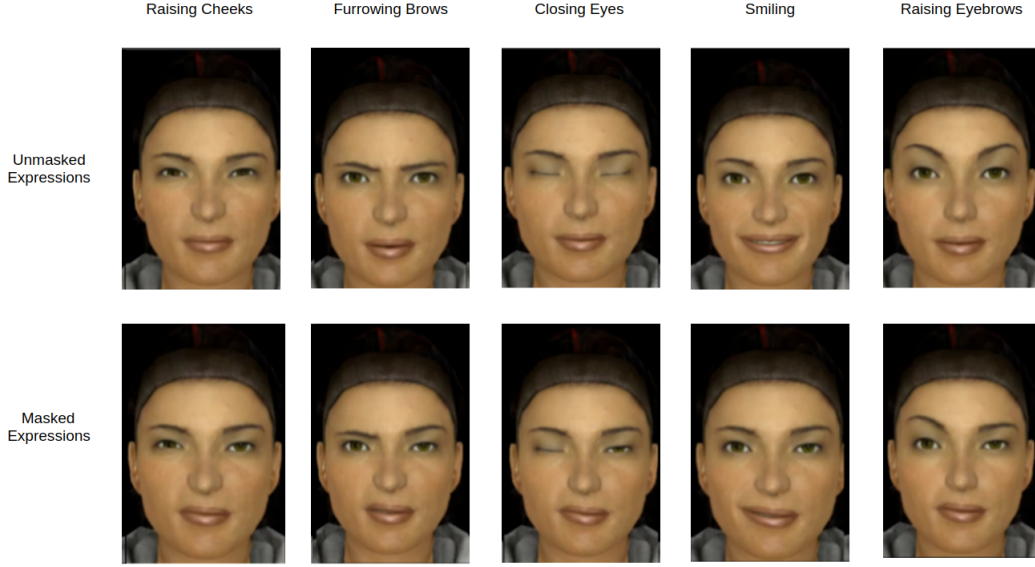


Fig. 4. Sample frames from the stimuli videos.

All the  $\beta_i$  scales are one for the 29 facial points on the unaffected side of the face. All the  $\beta_i$  scales are less than one for the 29 facial points on the affected side of the face. This is because FP causes weakness in the affected side of the face and therefore, feature points of that side will have less movement.

For each feature point on the affected side of the face, we calculated its scaling parameter in all of the frames and averaged them. For our work, we will consider frames with the highest intensity to build our model. Possible future work could be building a model of different pathologies in different stages of the disease with different intensities.

Additionally, applying the same method on different people with BP may output different scales and yield different models. This is because BP has some idiosyncratic characteristics; although all cases of BP include some degree of weakened mobility on one side of the face, the severity may vary across individuals. Thus, we built three masks using source videos of three people with BP (see Fig. 5).

#### IV. EXPERIMENTAL VALIDATION

To evaluate our masked synthesis module and get feedback for further refinement, we ran a qualitative, expert-based perceptual experiment. This is a common method for evaluating synthesized facial expressions [7], [24]. Getting initial feedback from clinicians is very valuable because it provides insights for improving the model and exploring its potential for being used as part of an educational tool for clinicians. Additionally, future users of our masked synthesis module are clinicians, and developing an educational tool for clinicians should be an interdisciplinary collaboration involving medical educators and computer scientists [15], [17], [18]. The experts in our study were four clinicians familiar with assessing facial paralysis in BP patients.

We first collected videos from a performer without BP. As ultimately we want this masked synthesis approach to be

used by operators without BP in a clinical simulation context (see Fig. 2), it was important to study the likely expressions clinicians would make, and how they would appear when masked. We recorded two sets of videos that are required in BP diagnosis: diagnostic expressions of the eyes, brow, and lower face, and expressions that happen during interviews between clinicians and BP patients.

In the first set (diagnostic), a performer without BP was recorded while performing five expressions required for assessing paralysis [16]: *closing the eyes*, *raising the eyebrows*, *furrowing the brow*, *smiling*, and *raising the cheeks*. The performer was instructed to repeat each expression five times. This resulted in five videos, each about 10 seconds long.

In the second set (interview), a performer without BP answered a list of questions that a clinician asks when diagnosing a BP patient [16]. Our goal was to record natural expressions that happen during interaction between a BP patient and a clinician. The interview video was 40 seconds long. This yielded five diagnostic video and one interview video.

From these six source videos, we used a face tracker [4] to track 68 facial points frame-by-frame, and extracted facial features of the performer’s face. We then applied the scaling parameters corresponding to each of the pre-built masks of BP to the extracted feature points of each source video. Since we had six source videos and three pre-built masks, this process gave us 18 sets of masked feature points.

For each set of masked feature points, we ran the Robot Operating System (ROS) module from [27] to synthesize the masked expressions to a virtual character’s control points for animation in Steam Source SDK. At the end of this step, we had 18 stimuli videos (15 diagnostic stimuli videos of expressions required for assessing BP, and three naturalistic stimuli videos).

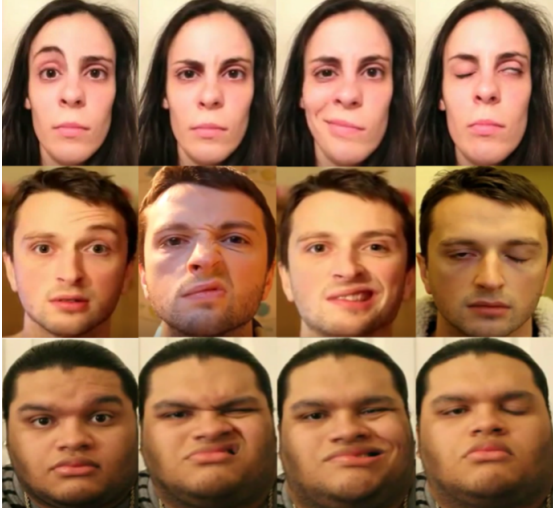


Fig. 5. Sample frames from three people with Bell’s palsy. We used these videos to develop our masks.

### A. Case Study

We sought feedback on the realism of our model when applied to a virtual character for synthesizing BP, and how similar the synthesized expressions are to those of real people with BP. We recruited four clinicians (one male and three female) from a US-based medical school. Participants were all native English speakers, and aged between 35 to 59 years old (mean age 48.7 years old). Three of the clinicians were physicians, and one was a clinical nurse. They all had on average 21 years of face-to-face interaction with patients and all had encountered BP patients in their careers.

We designed a structured online questionnaire to show them our stimuli videos, and ask for their feedback on various aspects of our synthesis results. At the beginning of the study, clinicians received instructions and a brief summary of the project. Next, they watched a video of a real person with BP performing various expressions. We showed them this video to refresh their memory on facial expressions of patients with BP. This video was not used in developing any of the three models.

We then asked participants questions in two parts. In the first part, they watched the 15 diagnostic stimuli videos in a random order. In each video, a cross-hair was shown for three seconds, followed by a virtual avatar conveying one of

the expressions required for assessing BP. They then rated the similarity of each avatar’s expressions to those of real BP patients for the entire face, and for that particular facial part (e.g. eyebrows for the raising eyebrow videos).

The similarity rating was a 4-point Discrete Visual Analogue Scale (DVAS). A one on the scale corresponded to “not at all similar to real patients” and a four on the scale corresponded to “very similar to real patients”. They could watch each video as many times as they need. Figure 4 shows sample frames from the first part of the study.

In the second part of the study, participants watched the three stimuli videos of naturalistic expressions in a random order. In each video, a cross-hair was shown for three seconds, followed by a virtual avatar conveying natural expressions that would happen during interaction with a clinician. Participants were then asked to rate the similarity of the avatar’s expressions to those of real Bell’s palsy patients overall (entire face), and for each facial part (eyes, eyebrows, cheeks, and mouth) on the same 4-point Discrete Visual Analogue Scale (DVAS).

Additionally, after watching each naturalistic video, they were asked to provide feedback by answering these open-ended questions: “Based on the video you just watched, what aspects did you think were the most realistic?” and “What areas did you think could be improved?”. After completing both parts of the study, clinicians were asked to answer these two open-ended questions: “How useful could this kind of simulation tool used in clinical education?” and “Please share any additional feedback or comments.”

## V. RESULTS

The overall average score for similarity between the synthesized masked expressions and real patients for the diagnostic expressions was 2.66 (s.d. =0.98). Table I reports the full results for each of the five diagnostic expressions and the entire face. This table suggests that overall clinicians thought that closing eyes and smiling were the most similar expressions to real patients. Cheeks had the lowest overall similarity scores, suggesting that we need to improve our model around the cheeks.

These similarity scores also show that overall the second mask created more realistic diagnostic expressions. For all the diagnostic expressions (closing eyes, smiling, furrowing brow, and raising cheeks), except raising eyebrows, the second mask outperformed (or was as good as) the other two

TABLE I  
MEAN SIMILARITY SCORES FOR EACH FACIAL PART AND THE ENTIRE FACE FOR THE DIAGNOSTIC STIMULI VIDEOS ON A 4-POINT DVAS.

	Mask 1		Mask 2		Mask 3		Overall	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
<b>Raising cheeks</b> (Cheeks)	1.75	0.95	<b>2.75</b>	0.95	1.75	0.50	2.08	0.90
<b>Furrowing brow</b> (Inner eyebrows)	<b>2.75</b>	1.25	<b>2.75</b>	0.95	<b>2.75</b>	1.50	<b>2.75</b>	1.13
<b>Closing eyes</b> (Eyes)	<b>3.25</b>	0.95	<b>3.25</b>	0.50	2.50	1.29	3.00	0.95
<b>Smiling</b> (Mouth)	2.50	0.57	<b>3.50</b>	0.57	3.00	1.15	3.00	0.85
<b>Raising eyebrows</b> (Outer eyebrows)	<b>2.75</b>	0.95	2.50	1.29	2.25	0.50	2.50	0.90
<b>Overall</b> (entire face)	2.60	0.99	<b>2.95</b>	0.88	2.45	1.05	2.66	0.98

TABLE II  
MEAN SIMILARITY SCORES FOR THE SYNTHESIZED INTERVIEW EXPRESSIONS ON A 4-POINT DVAS.

	Mask 1		Mask 2		Mask 3		Overall	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
<b>Cheeks</b>	2.00	1.00	2.00	1.00	<b>2.33</b>	0.57	2.11	0.78
<b>Eyebrows</b>	2.33	0.57	2.66	0.57	<b>3.00</b>	1.00	2.66	0.70
<b>Eyes</b>	<b>3.66</b>	0.57	3.33	0.57	3.33	1.15	3.44	0.72
<b>Mouth</b>	2.33	0.57	2.00	1.00	<b>2.66</b>	1.15	2.33	0.86
<b>Overall</b> (entire face)	2.66	0.57	2.66	0.57	<b>3.33</b>	0.57	2.88	0.60

models. This suggests that before using the second model in a larger study, its appearance around the eyebrows needs improvement.

For the naturalistic expression videos, we were only able to include the ratings from three of the clinicians as one was not able to participate in the second part of our study. Table II reports the full results for the naturalistic expressions. For the entire face, and all the facial parts except eyes, the third mask created natural expressions that are more similar to real BP patients. This table also suggests that overall clinicians thought that eyes were the most similar part of the face to real patients and cheeks were the least similar part of the face to real BP patients.

## VI. DISCUSSION

The work described in this paper is the first step towards building a complete, expressive patient simulator system capable of recognizing and synthesizing asymmetric facial expressions. Our work addresses the aforementioned gap in the literature by providing clinical learners with a training tool to practice their procedural and communication skills. In this paper, we introduced methods for modelling and synthesizing asymmetric facial expressions associated with BP, and experimentally validated our approach within a clinical education context.

Our experimental results are encouraging, and suggest that these techniques can be useful in affective computing, clinical education, and related fields. Our proposed algorithm generated a reliable computational model of BP. Moreover, our synthesis method is also suitable for use in real-time clinical simulation and training contexts. We were able to generate realistic face models conveying asymmetric facial expressions.

There are several ways to further enhance the models and techniques described. First, employing more control points can make the models more realistic. The avatar used in this work did not have any control points for creating wrinkles around the cheeks and nose. This led clinicians to select low ratings for the similarity of the expressions around the avatar's cheeks, when assessing both the diagnostic and the naturalistic expressions. Clinicians also suggested that adding control points for the eyes could increase the realism of the model, since the affected eyeball of people with BP often rolls into their head when they try to blink.

Second, we can explore including static asymmetry (e.g., facial drooping) to the synthesized expressions in future

experiments. Clinicians recommended adding forehead lines to create a more realistic furrowed brow expression, making it possible to recognize the intensity of BP. Moreover, they mentioned that adding "puffing out the cheeks" to the diagnostic set can also improve the expressions, as they are useful for assessing asymmetry of the mouth and lips.

Third, ideally, we can create a general model for BP that encompasses all the predominant features of BP in order to generally represent all BP patients. If we have enough input source videos of people with BP, we may be able to extract the common features of this specific pathology to create one general BP model. Similarly, if datasets for other pathologies become publicly available, we could also extend this technique to them (e.g., stroke, dystonia).

Finally, we ultimately want to run a similar experiment with clinical learners using an expressive RPS. This may help clinicians avoid forming biased impressions of people with FP and improve their cultural competency. Also, it may improve their skills in understanding the emotions of people with asymmetric facial expressions, which can lead to better health outcomes and reduce health disparities.

Our work is important for the affective computing and robotics communities because it allows researchers to explore new methods of facial expression synthesis. For example, our work can be used in teleconferencing or animation to control and manipulate the expressions and identity of a virtual character. It can also be used in computer games, to mask the identity or expressions of a user, or in social assistive robots to be used to support health and well being.

Overall, this work makes contributions to both interactions with virtual humans, and the study of clinical conditions that alter expressive behaviour, which are two critical applications of facial expression analysis [41]. More specifically, the exploration of recognizing and synthesizing asymmetric facial expressions have significant implications for human emotion analysis.

## ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant No. IIS-1253935.

## REFERENCES

- [1] *The world health report 2002: reducing risks, promoting healthy life*. World Health Organization, 2002.
- [2] R. C. Arkin and M. J. Pettinati. Moral emotions, robots, and their role in managing stigma in early stage parkinsons disease caregiving. In *Workshop on New Frontiers of Service Robotics for the Elderly, RO-MAN*, 2014.

- [3] T. Baltrušaitis, P. Robinson, and L. Morency. 3d constrained local model for rigid and non-rigid facial tracking. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.
- [4] T. Baltrušaitis, P. Robinson, and L. Morency. Openface: an open source facial behavior analysis toolkit. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2016.
- [5] A. Bandini, J. Green, B. Richburg, and Y. Yunusova. Automatic detection of orofacial impairment in stroke. In *Proc. Interspeech 2018*, 2018.
- [6] R. F. Baugh, G. J. Basura, L. E. Ishii, S. R. Schwartz, C.M. Drumheller, and R. Burkholder. Clinical practice guideline bells palsy. *Otolaryngology–Head and Neck Surgery*, 2013.
- [7] C. Becker-Asano and H. Ishiguro. Evaluating facial displays of emotion for the android robot Geminoid F. In *IEEE Workshop on Affective Computational Intelligence (WACI)*, 2011.
- [8] M. Benz, X. Laboureaux, T. Maier, E. Nkenke, S. Seeger, F. Neukam, and G. Husler. The symmetry of faces. 2002.
- [9] K. R. Bogart and L. Tickle-Degnen. Looking beyond the face: A training to improve perceivers impressions of people with facial paralysis. *Patient education and counseling*, 2015.
- [10] K. R. Bogart, L. Tickle-Degnen, and N. Ambady. Communicating without the face: Holistic perception of emotions of people with facial paralysis. *Basic and Applied Social Psychology*, 2014.
- [11] S. M. Boker, J. F. Cohn, B. Theobald, I. Matthews, T. R. Brick, and J. R. Spies. Effects of damping head movement and facial expression in dyadic conversation using real-time facial expression tracking and synthesized avatars. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 2009.
- [12] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Trans. Pattern Anal. Mach. Intell.*, 1998.
- [13] D. Cristinacce and T. F. Cootes. Feature detection and tracking with constrained local models. *Pattern Recognition*, 2006.
- [14] H. Dai, N. Pears, W. Smith, and C. Duncan. Symmetric shape morphing for 3d face and head modelling. In *2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)*, 2018.
- [15] A. M. Deladisma, M. Cohen, A. Stevens, P. Wagner, B. Lok, T. Bernard, C. Oxendine, L. Schumacher, K. Johnsen, R. Dickerson, et al. Do medical students respond empathetically to a virtual patient? *The American Journal of Surgery*, 2007.
- [16] T. J. Eviston, G. R. Croxson, P. G. Kennedy, T. Hadlock, and A. V. Krishnan. Bell's palsy: aetiology, clinical features and multidisciplinary care. *Journal of Neurology, Neurosurgery & Psychiatry*, 2015.
- [17] M. J. Gonzales, V. C. Cheung, and L. D. Riek. Designing collaborative healthcare technology for the acute care workflow. In *Proceedings of the 9th International Conference on Pervasive Computing Technologies for Healthcare*, 2015.
- [18] M. J. Gonzales, J. M. Henry, A. W. Calhoun, and L. D. Riek. Visual TASK: A collaborative cognitive aid for acute care resuscitation. *10th EAI International Conference on Pervasive Computing Technologies for Healthcare (Pervasive Health)*, 2016.
- [19] J. T. James. A new evidence-based estimate of patient harms associated with hospital care. *Journal of patient safety*, 2013.
- [20] A. Kalegina, G. Schroeder, A. Allchin, K. Berlin, and M. Cakmak. Characterizing the design space of rendered robot faces. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, 2018.
- [21] Y. Kihara, G. Duan, T. Nishida, N. Matsushiro, and Y. Chen. A dynamic facial expression database for quantitative analysis of facial paralysis. In *2011 6th International Conference on Computer Sciences and Convergence Information Technology (ICCIT)*, 2011.
- [22] A. G. Marson and R. Salinas. Clinical evidence: Bell's palsy. *Western Journal of Medicine*, 2000.
- [23] T. J. Martin, A. P. Rzepczynski, and L. D. Riek. Ask, inform, or act: communication with a robotic patient before haptic action. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, 2012.
- [24] D. Mazzei, N. Lazzeri, D. Hanson, and D. De Rossi. Hefes: An hybrid engine for facial expressions synthesis to control human-like androids and avatars. In *IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, 2012.
- [25] M. Moosaei, S. K. Das, D. O. Popa, and L. D. Riek. Using facially expressive robots to calibrate clinical pain perception. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction (HRI 2017)*, 2017.
- [26] M. Moosaei, M. J. Gonzales, and L. D. Riek. Naturalistic pain synthesis for virtual patients. In *International Conference on Intelligent Virtual Agents*, 2014.
- [27] M. Moosaei, C. J. Hayes, and L. D. Riek. Facial expression synthesis on robots: An ros module. *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, 2015.
- [28] M. Moosaei, C. J. Hayes, and L. D. Riek. Performing facial expression synthesis on robot faces: A real-time software system. In *Proceedings of the 4th International AISB Symposium on New Frontiers in Human-Robot Interaction, AISB*, 2015.
- [29] P. Morton. Creating a laboratory that simulates the critical care environment. *Critical Care Nursing*, 16(6), 7681. *National Association for Associate Degree Nursing*, 1995.
- [30] T. H. Ngo, Y. Chen, M. Seo, N. Matsushiro, and W. Xiong. Quantitative analysis of facial paralysis based on three-dimensional features. In *2016 IEEE International Conference on Image Processing (ICIP)*, 2016.
- [31] T. H. Ngo, M. Seo, Y. Chen, and N. Matsushiro. Quantitative evaluation of facial paralysis using tracking method. In *The 2015 IEEE RIVF International Conference on Computing Communication Technologies - Research, Innovation, and Vision for Future (RIVF)*, 2015.
- [32] M. Pourebadi and L. D. Riek. Expressive robotic patient simulators for clinical education. In *R4L Workshop on Robots for Learning - Inclusive Learning, Workshop at the 13th Annual ACM/IEEE International Conference on Human Robot Interaction (HRI 2018)*, 2018.
- [33] L. D. Riek. Robotics technology in mental healthcare. *D. Luxton (Ed.), Artificial Intelligence in Behavioral Health and Mental Health Care. Elsevier*, pp. 185-203., 2015.
- [34] L. D. Riek. Healthcare robotics. *Communications of the ACM*, Vol. 60, No. 11, pp. 68-78, 2017.
- [35] L. D. Riek and P. Robinson. Using robots to help people habituate to visible disabilities. In *IEEE International Conference on Rehabilitation Robotics (ICORR)*, 2011.
- [36] K. L. Robey, P. M. Miniham, L. M. Long-Bellil, J. E. Hahn, J. G. Reiss, G. E. Eddey, Alliance for Disability in Health Care Education, et al. Teaching health care students about disability within a cultural competency context. *Disability and Health Journal*, 2013.
- [37] H. Sunvisson, B. Habermann, S. Weiss, and P. Benner. Augmenting the cartesian medical discourse with an understanding of the person's lifeworld, lived body, life story and social identity. *Nursing Philosophy*, 2009.
- [38] L. Tickle-Degnen, M. Scheutz, and R. C. Arkin. Collaborative robots in rehabilitation for social self-management of health. 2014.
- [39] L. Tickle-Degnen, L. A. Zebrowitz, and H. Ma. Culture, gender and health care stigma: Practitioners response to facial masking experienced by people with parkinsons disease. *Social Science & Medicine*, 2011.
- [40] O. Tysnes and A. Storstein. Epidemiology of parkinson's disease. *Journal of Neural Transmission*, 2017.
- [41] M. Valstar. Automatic behaviour understanding in medicine. In *Proceedings of the 2014 Workshop on Roadmapping the Future of Multimodal Interaction Research Including Business Opportunities and Challenges*, 2014.
- [42] S. Wang and L. Tickle-Degnen. Emotional cues from expressive behavior of women and men with parkinsons disease. *PloS one*, 2018.