INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

Toward Development and Evaluation of Pain Level-Rating Scale for Emergency Triage based on Vocal Characteristics and Facial Expressions

Fu-Sheng Tsai[1], *Ya-Ling Hsu*[1], *Wei-Chen Chen*[1], *Yi-Ming Weng*[2], *Chip-Jin Ng*[2], *Chi-Chun Lee*[1]
[1]Department of Electrical Engineering, National Tsing Hua University, Taiwan
[2]Department of Emergency Medicine, Chang Gung Memorial Hospital, Taiwan

Abstract

In order to allocate the healthcare resource, triage classification system plays an important role in assessing the severity of illness of the boarding patient at emergency department. The self-report pain intensity numerical-rating scale (NRS) is one of the major modifiers of the current triage system based on the Taiwan Triage and Acuity Scale (TTAS). The validity and reliability of *self-report* scheme for pain level assessment is a major concern. In this study, we model the observed expressive behaviors, i.e., facial expressions and vocal characteristics, directly from audio-video recordings in order to measure pain level for patients during triage. This work demonstrates a feasible model, which achieves an accuracy of 72.3% and 51.6% in a binary and ternary pain intensity classification. Moreover, the study result reveals a significant association of current model and analgesic prescription/patient disposition after adjusted for patient-report NRS and triage vital signs.

Index Terms: behavioral signal processing (BSP), facial expressions, triage, pain scale, vocal characteristics

1. Introduction

Deriving behavioral informatics from signals, e.g., audio-video and/or physiological data recordings, offers a new paradigm for quantitative decision-making across behavior sciences [1]. Behavioral informatics, i.e., computational methods that measure human's attributes-of-interest, are developed grounded in their desired domain applications. For example, notable algorithmic advances have been observed in the medical domains: detection of depression [2, 3], assessment of Parkinson's disease [4, 5], modeling of therapist's empathy in motivational interview [6, 7], analysis of disorder [8, 9], etc. In this work, we carry out a research effort into objectifying pain level, i.e., one of the six major regulators in the Taiwan Triage and Acuity Scale (TTAS) [10], of an on-boarding emergency patient by modeling his/her facial expressions and vocal characteristics.

TTAS is jointly developed by the Taiwan Society of Emergency Medicine and the Critical Care Society, which modifies the Canadian Triage and Acuity Scale (CTAS) [11] by tailoring toward Taiwan's particular medical situations. It is officially announced in 2010 by the Ministry of Health and Welfare to be the triage system of Taiwan. TTAS includes six major factors in assessing the severity and screening life-threatening patients: respiratory distress, circulation, consciousness level, body temperature, pain level, and injury mechanism. In specifics, the intensity of pain is currently measured by the numerical-rating scale (NRS)[12, 13], that is a 10-point self-report pain scale. In clinical practice, physicians and nurses have noticed the difficulty in the systematic implementation of this instrument especially for elderly people, foreigners, or patients with a low education level. This often leads to either a practice of using FACES rating scale [14] that is designed for children, or the triage nurses would select the level through his/her own observations instead of soliciting an answer from the patient. Furthermore, even when the nurses succeed in carrying out NRS, this *self-report* rating still suffers from various unwanted idiosyncratic factors, e.g., age and body part dependency and inconsistent comprehension of the pain scale. These issues centered around subjectivity in measuring pain create a deviation on the consistency and validity of the triage classification system.

Related previous works have concentrated mainly on recognizing the occurrences of pain by monitoring facial expressions. For example Ashraf *et al.* [15] uses active appearance model to recognize frame-level pain, Kaltwang *et al.* [16] uses a relevance vector regression model to classify between real and fake pain, and Wener *et al.* models head pose for pain detection [17]. In this work, we propose to include voice characteristics in addition to facial expression for measuring pain level. Moreover, we contribute not only in the multimodal aspect of pain level measurement but also in the realism of *contextualizing* applications in real medical settings. In this work, we collect data from a total of 182 real patients as they seek emergency medical service at Chang Gung Memorial Hospital[1]. The data includes audio-video samples during triage and follow-up sessions after treatment, vital sign (physiological) data during triage, and finally a set of clinical outcomes. The data recordings are in real medical settings (in-the-wild), and the interactions are spontaneous in nature - all posed as a challenging yet contextualized situation for deriving appropriate informatics.

Our proposed multimodal framework achieves a 72.3% accuracy in classifying between the extremes (*severe* versus *mild*) pain level and 51.6% accuracy in performing a three-class (*severe*, *moderate*, and *mild*) pain level recognition. The inclusion of audio modality is essential in improving the overall recognition rate implicating that the intensity of pain is also reflected in the patient's vocal characteristics. Furthermore, while comparing to the so called ground-truth, i.e., NRS, is a straightforward mean of evaluating the framework, in this work, we further utilize this audio-video based system in combination with NRS and vital sign data in order to analyze clinically-relevant outcome-related variables, in specifics analgesic prescription and patient disposition, as another evaluation scheme. We demonstrate that even after taking into the account for the current best medical instruments availability (physiological data and NRS), the usage of audio-video based pain level assessment can improve the prediction about whether a doctor would end up prescribing analgesic prescription or order patients to be hospitalized. This initial result is quite promising as the research effort will continue to derive novel pain-level rating and validate its ability to improve the current triage classification system clinically.

[1]IRB#:104-3625B

http://dx.doi.org/10.21437/Interspeech.2016-408

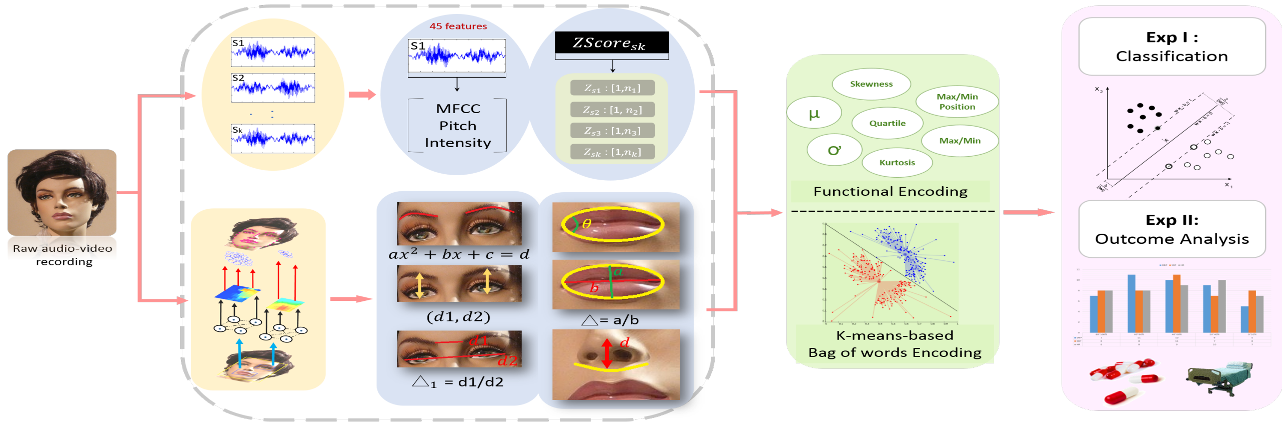

Figure 1: *It shows a complete flow diagram of the proposed work. We segment the raw audio recordings manually and then extract acoustic low-level descriptors; for the video data, we apply a pre-trained constraint local neural field (CLNF) to track the* (x, y) *positions of the 68 landmark points on face and then extract descriptors based on to pain-related facial action units. Two encodings methods, statistical functional descriptors and k-means bag-of-word model, are used to derive a session-level feature vector. Finally, we conduct pain level recognition using fusion of audio-video features and further analyze it with respect to clinical outcomes.*

The rest of the paper is organized as follows: section 2 describes about data collection and audio-video feature extraction, section 3 includes experimental setups and results, and section 4 concludes with future work.

2. Research Methodology

2.1. Database Collection

The triage session included audio-video recordings, physiological (heart rate, systolic and diastolic blood pressure) vital sign data, and other clinically-related outcomes (analgesic prescription and patient disposition) of on-boarding emergency patients at Chang Gung Memorial Hospital. We excluded pediatric and trauma patients, also excluded referral patients or patients with prior treatment before arrival, and further only included patients with symptoms of chest, abdominal, lower-back, limbic pain, and headaches. There were two sessions recorded for each patient, i.e., at triage and follow-up, where the follow-up session occurred approximately 1 hour after the treatment, if any, was given to the patient. These sessions essentially involved nurses asking the patient for the location of the body pain, the NRS scale of pain intensity ($0 - 10$, where 10 means the worst pain ever), and a brief description on the type of pain felt (for example, cramps or aches); it usually lasted around 30 seconds for each session. The audio-video data was recorded using a Sony HDR handy cam on a tripod in a designated assessment room, and the placement of the camera was set attempting to consistently capture the patients' facial expressions.

In our current database, we have collected a total of 182 patients, each recorded at the two designated points in time. After excluding non-usable data (e.g., cases where the patient's relative responds to the pain level assessment instead of the patient, low audio-video quality due to various uncontrollable factors, loss of either physiological data or clinical outcomes), we have a total of 205 audio-video samples from 117 unique patients, which constitutes the dataset of interest for this work. Lastly, the pain level is often grouped into three levels based on the number reported (mild: $0-3$, moderate: $4-6$, severe: $7-10$); we adopt the same convention in this work to serve as the learning target for our signal-based pain level assessment system.

2.2. Audio-Video Feature Extraction

Figure 1 depicts the overall framework including audio-video data preprocessing, low-level descriptors extraction, and session-level encoding. In the following sections, we will briefly describe each component.

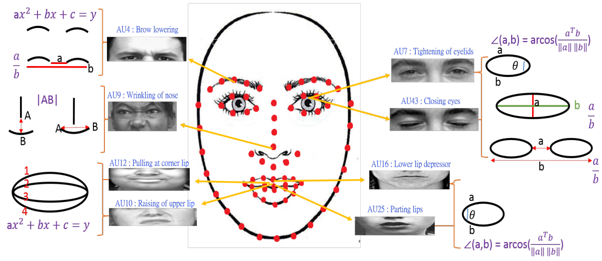

Figure 2: *The red dots are the 68 facial landmarks tacked for each image. The action units are the ones being indicative of pain in the past literature. Lastly, it shows the various parameterization of the 68 facial landmarks that we compute as video features to be used in this work. Facial Action Coding System photos(http://www.cs.cmu.edu/ face/facs.htm)*

2.2.1. Acoustic Characteristics

For each recorded session, we first perform manual segmentation on the audio file to obtain the speaking portions corresponding to the patient, the patient's relatives, and the interviewer. In this work, we concentrate only on the patient's voice characteristics. We extract 45 low-level descriptors in total, including 13 MFCCs, 1 fundamental frequency, 1 intensity and their associated delta and delta-delta every 10ms. This set of spectral-prosodic features is extracted due to their common usage in characterizing paralinguistic and emotion information [18]. The audio features are further z-normalized per speaker.

2.2.2. Facial Expressions

On the video side, for each session, we first apply constrained local neural fields (CLNF) [19] as a pre-processing step. CLNF tracks a patient's 68 facial landmark's position based on the Active Orientation Model (AOM) [20], which is an extension to Active Appearance Model for describing the shape and appearance of a face. CLNF, i.e., an instance of constrained local model, essentially involves three major technical components: point distribution model (describing the position of feature points in an image), local neural field patch experts (layered unidirectional graphical model), and optimization fitting approach (non-uniform regularized landmark mean shift fitting). By applying CLNF, we then are able to track the 68 feature points (Figure 2), e.g., around face, eyes, and nose contour, for each patient in each image of the recorded video session.

Past works have identified several facial action units that are related to the feeling of pain [21, 22], e.g., AU4, 6, 7, 9, 10, 12,

Table 1: *It summarizes the Unweighted Average Recall (UAR) obtained in Exp I. 2-Class indicates the binary classification task between the extreme pain levels (severe versus mild). 3-Class indicates the ternary classification between severe, moderate, and mild pain levels. The numbers in bold indicate the best accuracy achieved within that specific task.*

	Chance	Audio-Only		Video-Only		Multimodal Fusion (*early*-fusion / *late*-fusion)			
		Functional	**Bow**	**Functional**	**Bow**	**FuncA, FuncV**	**FuncA, BowV**	**BowA,FuncV**	**BowA, BowV**
2-Class	50.0	67.9	61.3	55.9	61.9	66.8 / 68.7	**72.3** / 68.1	56.6 / 61.5	61.1 / 64.8
3-Class	33.3	46.0	42.9	40.9	40.8	43.5 / 48.3	49.7 / **51.6**	40.8 / 43.7	41.4 / 44.8

16, 25, 43 (Figure 2). In this work, instead of recognizing these facial action units, we compute features characterizing these expressions directly from the tracked key points' (x, y) position,

- *Eyebrows* (7): the distance of inner eyebrows divided by the distance of outer eyebrows (1), the quadratic polynomial coefficients of the right and the left eyebrows (6)
- *Nose* (2): the normalized distance between nose and philtrum (1), and of nasolabial folds (1)
- *Eyes* (5): the outer eye corner opening (2), the distance between the inner eye corners divided by the distance of outer eye corners (1), the distance of upper and lower eyelids divided by the distance from the head to the corner of the eyes (2)
- *Mouth* (14): the quadratic polynomial coefficients from the shape of upper lip, including outer and inner part, and lower lip, including outer and inner part (12), the two-sided mouth corners opening angles (2)

There are a total of 28 features per frame extracted from the face to represent the facial expression of the patient. Figure 2 also shows a schematics of the features being extracted in this work.

2.3. Session-level Encodings

Since each session is approximately 30 seconds long, we additionally utilize two different encoding approaches to form a fixed-length feature vector at the session-level. The first one is based on computing 15 different statistical functionals on audio and video low-level descriptors (*Functional*). The list of functionals includes maximum, minimum, mean, median, standard deviation, 1st percentile, 99th percentile, 99th-1st percentile, skewness, kurtosis, minimum position, maximum position, lower quartile, upper quartile, and interquartile range. The second approach is based on k-means bag-of-word (*BoW*) encoding, which encoding varying length of sequences of low-level descriptors with a histogram count of cluster occurrences. In general, BoW characterizes the quantified behavior types over a duration of time. The number of clusters is set to be 256 for both audio and video.

3. Experimental Setup and Results

In this work, we set up two different experiments:

- **Exp I** : the NRS pain scale recognition task
- **Exp II**: clinical outcomes analyses

Exp I is designed to validate that the pain-related facial and vocal expressions can indeed be modeled and used in the development toward a signal-based pain scale, and Exp II is designed to analyze the predictive information that the signal-based pain scale possess in addition to the NRS and patient's physiology to the clinical judgment of painkiller prescription and patient's disposition (hospitalization or discharge).

3.1. Exp I: NRS Pain Level Classification

In Exp I, we perform two different recognition tasks: 1) binary classification between the extreme pain levels, i.e., severe vs. mild pain, on the subset of the dataset and 2) ternary classification of the three commonly-used pain levels, i.e., severe vs. moderate vs. mild, on the entire dataset. Severe pain corresponds to NRS score ranging between $7-10$, moderate is $4-6$, and mild is $0-3$. We design two different tasks due to the fact that NRS rating itself only relies on patient's *self-report*, which can be subjective especially for the *moderate* portion of the data. By running an additional binary classification on the extreme set, where there is less concern on the reliability of the label, we can better assess the technical feasibility of our framework. The classifier of choice for this experiment is the linear-kernel support vector machine. We employ two different multimodal fusion techniques. One is based on *early*-fusion technique, i.e., concatenating audio and video features after performing univariate feature selection (i.e., ANOVA) on each modality separately. Another one is based on *late*-fusion technique, i.e., by fusing the decision scores from the audio and video modality separately using logistics regression. All evaluation is done via leave-one-patient-out cross-validation, and the performance metric is unweighted average recall.

3.1.1. Results and Discussions

Table 1 summarizes the results of Exp I. *2-Class* indicates the binary classification task between the extremes. *3-Class* indicates the ternary classification between the three pain levels. The numbers in bold indicate the best accuracy achieved. There are a couple of points to note in these results. The best accuracies achieved are 72.3% and 51.6%, i.e., multimodal fusion of audio and video modalities, for *2-Class* and *3-Class* classification tasks respectively. Both of these results are significantly better than the chance baseline indicating that there indeed exists pain-related information that can be modeled through audio-video signals. Another point to make is that while past works concentrate mostly on the facial expressions, in our work, we demonstrate that the vocal characteristics are also indicative of the patient's experience of pain. In fact, if we compare the audio-only and video-only accuracies, the result obtained with audio-only features are slightly higher than the video-only features.

Secondly, the type of encoding methods affects the recognition accuracies. We show that functionals-based method works better for audio features and bag-of-word approach works better for video features. In fact, the best accuracy reported is by fusing functional-based audio feature with bag-of-word encoding of video feature. We hypothesize that this could be due to the fact that pain-related audio characteristics are non-linearly distributed across the session (hence, the functional descriptor approach works better), and our video features are inherently trying to capture a specific configuration of appearances (hence, a counting-based method of encoding is superior). Another thing to note that, in the three-class problem, the error rate for the *moderate* class is considerably higher than in the *mild* and *severe*. It could be due to the fact that this class is inherently ambiguous; hence not only the data itself is ambiguous but the ground truth itself can be unreliable. In summary, we demonstrate that our proposed audio-video-based pain scale is capable of reaching a substantial reliability compared to the established NRS self-report-based instrument for assessing pain.

3.2. Exp II: Clinical Outcomes Analyses

The overarching goal of the research effort is not just to *replicate* the NRS self-report pain scale, instead, the aim is to derive a signal-based (i.e., from audio-video data) informatics that can supplement the current decision-making protocol. A physician's decision on the type of treatment, if any, to the patient is often largely based on a holistic clinical assessment of a patient's overall condition. Hence, in Exp II, our aim is to design a simple quantitative *score* that combines the available measures at triage with the audio-video based pain level (system output in section 3.1). We will demonstrate that this score has *added* information that is relevant to the patient's clinical outcomes of analgesic prescription and disposition. The exact analysis procedure goes as follows. For each triage, we have the following measures for every patient:

- **PHY**: age, systolic/diastolic blood pressure, heart rate
- **NRS-3C**: the three pain levels, i.e., mild, moderate, and severe, derived from the patient's NRS scale
- **SYS-2C**: one of the two predicted pain levels (mild / severe) derived from the *2-Class* SVM
- **SYS-2C(d)**: the decision score derived the *2-Class* SVM
- **SYS-3C**: one of the three predicted pain levels (mild / moderate / severe) derived from the *3-Class* SVM
- **SYS-3C(d)**: the decision score derived the *3-Class* SVM

PHY measures are all normalized with respect to the age of each patient. Further, we have two clinical dichotomous outcome variables for each patient, i.e., painkiller prescription and disposition. We design a score, *painK* and *dispT*, of each outcome by training a linear regression model each on the training set using the measures mentioned above as the independent variables, and then we apply the learned regression model to assign an outcome score for each patient i. Lastly, by utilizing the following simple rule, we can predict whether a patient i will end up being prescribed medication or being hospitalized:

$$\text{prescription: painK}_i > \text{AVG}\{\text{painK}_j\}_{\forall j \in \text{train-set}}$$
$$\text{hospitalization: dispT}_i > \text{AVG}\{\text{dispT}_j\}_{\forall j \in \text{train-set}}$$

where AVG means the average values of the score within the training set. All of these procedures are done completely via leave-one-patient-out cross validation. The main idea of the analyses is to show that by having the audio-video based pain-scale system, it enhances the quantitative (i.e., objective and measurable) evidences to the doctor's clinical judgment even when accounted for the current clinical instruments.

3.2.1. Experimental Results and Discussions

Table 2 summarizes the results of Exp II as measured in UAR. There are some interesting points to note in this analysis. For the outcome of analgesic prescription, we see that NRS scale by itself naturally is already capable of achieving an accuracy of 66.3%, in accordance with known phenomenon in the past [23], and PHY measures alone do not contribute at all. However, by combing NRS to the SYS-3C(d), the accuracy improves to 71.0% (an 4.7% absolute improvement). This result seems to indicate that the decision scores outputted from the *3-Class* SVM encodes additional information beyond NRS scale that is relevant in understanding how physicians make a judgment on analgesic prescription. Furthermore, for the outcome of patient's disposition (hospitalization or not), we see that PHY (vital sign) by itself obtains 56.4% accuracy, where NRS scale does not provide information here. However, by coming PHY with SYS-2C, the accuracy improves to 65.7% (a 9.3% absolute improvement) - signifying the added information that NRS is lacking originally yet the audio-video based pain-scale do possess in terms patient's disposition outcome.

Table 2: *Summary of Exp II: the accuracy number is measured in unweighted average recall*

	Analgesic Pres.	Hospitalization
PHY	49.6	56.4
NRS-3C	66.3	42.7
PHY+NRS-3C	63.5	56.4
SYS-2C	51.5	58.7
SYS-3C	58.8	57.1
PHY+SYS-2C	47.8	**65.7**
PHY+SYS-3C	53.3	58.6
PHY+SYS-2C(d)	54.4	56.4
PHY+SYS-3C(d)	58.4	54.4
NRS-3C+SYS-2C	66.3	58.7
NRS-3C+SYS-3C	66.3	57.1
NRS-3C+SYS-2C(d)	66.3	43.3
NRS-3C+SYS-3C(d)	**71.0**	44.7
PHY+NRS-3C+SYS-2C	62.3	**65.7**
PHY+NRS-3C+SYS-3C	62.7	55.9
PHY+NRS-3C+SYS-2C(d)	66.0	55.1
PHY+NRS-3C+SYS-3C(d)	69.6	55.8

In summary, while the audio-video based system is trained from the NRS, it seems to differ possibly due to the fact that it models the facial expressions and vocal characteristics directly. We demonstrate that these signal-based pain scales indeed possess additional clinically-relevant information to the outcome variables of emergency triage beyond what is already captured in the NRS scale and conventional vital sign measures.

4. Conclusions

In this work, we develop an initial predictive framework to assess the pain-level for patients at emergency triage. The systems show reliable estimates to the established NRS pain scale. Furthermore, we evaluate the usefulness of such system by demonstrating that it can capture important information about the outcome of the patient beyond the current available instrumentations used at triage. This initial result is quite promising as the goal of the research is to devise a novel objective and quantifiable informatics not to replicate the current instrumentation but to provide supplemental clinically-relevant information that is beyond the established protocols.

There are multiple future directions. Technically, employing state-of-the-art speech/video processing and machine learning algorithms will be an immediate future direction as we continue to collect more data samples (our aim is to collect at least 500 unique patients' data). On the analysis part, we will put effort into understanding *exactly* what additional information that the system is able to capture from the facial and vocal expressions about the pain that is missing from the NRS scale, and whether such information is related to the physiology of the patient (e.g., muscle movement in response to pain felt that may correlate with the measures of heart rate or blood pressure). Having more insights discovered, we can hopefully help advance and benefit the current medical practices at the emergency triage with the introduction of such an informatics.

5. Acknowledgments

Thanks to MOST (103-2218-E-007-012-MY3) and ChangGung Memorial Hospital (CMRPG3E1791) for funding.

6. References

[1] S. Narayanan and P. G. Georgiou, "Behavioral signal processing: Deriving human behavioral informatics from speech and language," *Proceedings of the IEEE*, vol. 101, no. 5, pp. 1203–1233, 2013.

[2] J. F. Cohn, T. S. Kruez, I. Matthews, Y. Yang, M. H. Nguyen, M. T. Padilla, F. Zhou, and F. D. La Torre, "Detecting depression from facial actions and vocal prosody," in *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*. IEEE, 2009, pp. 1–7.

[3] Z. Liu, B. Hu, L. Yan, T. Wang, F. Liu, X. Li, and H. Kang, "Detection of depression in speech," in *Affective Computing and Intelligent Interaction (ACII), 2015 International Conference on*. IEEE, 2015, pp. 743–747.

[4] A. Tsanas, M. A. Little, C. Fox, and L. O. Ramig, "Objective automatic assessment of rehabilitative speech treatment in parkinson's disease," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 22, no. 1, pp. 181–190, 2014.

[5] A. Bayestehtashk, M. Asgari, I. Shafran, and J. McNames, "Fully automated assessment of the severity of parkinson's disease from speech," *Computer speech & language*, vol. 29, no. 1, pp. 172–185, 2015.

[6] J. Gibson, N. Malandrakis, F. Romero, D. C. Atkins, and S. Narayanan, "Predicting therapist empathy in motivational interviews using language features inspired by psycholinguistic norms," in *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.

[7] B. Xiao, D. Can, P. G. Georgiou, D. Atkins, and S. S. Narayanan, "Analyzing the language of therapist empathy in motivational interview based psychotherapy," in *Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012 Asia-Pacific*. IEEE, 2012, pp. 1–4.

[8] J. Kim, N. Kumar, A. Tsiartas, M. Li, and S. S. Narayanan, "Automatic intelligibility classification of sentence-level pathological speech," *Computer speech & language*, vol. 29, no. 1, pp. 132–144, 2015.

[9] D. Bone, C.-C. Lee, M. P. Black, M. E. Williams, S. Lee, P. Levitt, and S. Narayanan, "The psychologist as an interlocutor in autism spectrum disorder assessment: Insights from a study of spontaneous prosody," *Journal of Speech, Language, and Hearing Research*, vol. 57, no. 4, pp. 1162–1177, 2014.

[10] C.-J. Ng, Z.-S. Yen, J. C.-H. Tsai, L. C. Chen, S. J. Lin, Y. Y. Sang, J.-C. Chen *et al.*, "Validation of the taiwan triage and acuity scale: a new computerised five-level triage system," *Emergency Medicine Journal*, vol. 28, no. 12, pp. 1026–1031, 2011.

[11] M. J. Bullard, T. Chan, C. Brayman, D. Warren, E. Musgrave, B. Unger *et al.*, "Revisions to the canadian emergency department triage and acuity scale (ctas) guidelines," *CJEM*, vol. 16, no. 06, pp. 485–489, 2014.

[12] K. Eriksson, L. Wikström, K. Årestedt, B. Fridlund, and A. Broström, "Numeric rating scale: patients' perceptions of its use in postoperative pain assessments," *Applied nursing research*, vol. 27, no. 1, pp. 41–46, 2014.

[13] E. Castarlenas, E. Sánchez-Rodríguez, R. de la Vega, R. Roset, and J. Miró, "Agreement between verbal and electronic versions of the numerical rating scale (nrs-11) when used to assess pain intensity in adolescents," *The Clinical journal of pain*, vol. 31, no. 3, pp. 229–234, 2015.

[14] G. Garra, A. J. Singer, B. R. Taira, J. Chohan, H. Cardoz, E. Chisena, and H. C. Thode, "Validation of the wong-baker faces pain rating scale in pediatric emergency department patients," *Academic Emergency Medicine*, vol. 17, no. 1, pp. 50–54, 2010.

[15] A. B. Ashraf, S. Lucey, J. F. Cohn, T. Chen, Z. Ambadar, K. M. Prkachin, and P. E. Solomon, "The painful face–pain expression recognition using active appearance models," *Image and vision computing*, vol. 27, no. 12, pp. 1788–1796, 2009.

[16] S. Kaltwang, O. Rudovic, and M. Pantic, "Continuous pain intensity estimation from facial expressions," in *Advances in Visual Computing*. Springer, 2012, pp. 368–377.

[17] P. Werner, A. Al-Hamadi, R. Niese, S. Walter, S. Gruss, and H. C. Traue, "Towards pain monitoring: Facial expression, head pose, a new database, an automatic system and remaining challenges," in *Proceedings of the British Machine Vision Conference*, 2013, pp. 119–1.

[18] B. Schuller, S. Steidl, A. Batliner, F. Burkhardt, L. Devillers, C. MüLler, and S. Narayanan, "Paralinguistics in speech and languagestate-of-the-art and the challenge," *Computer Speech & Language*, vol. 27, no. 1, pp. 4–39, 2013.

[19] T. Baltrusaitis, P. Robinson, and L.-P. Morency, "Constrained local neural fields for robust facial landmark detection in the wild," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2013, pp. 354–361.

[20] G. Tzimiropoulos, J. Alabort-i Medina, S. Zafeiriou, and M. Pantic, "Generic active appearance models revisited," in *Computer Vision–ACCV 2012*. Springer, 2012, pp. 650–663.

[21] P. Lucey, J. F. Cohn, I. Matthews, S. Lucey, S. Sridharan, J. Howlett, and K. M. Prkachin, "Automatically detecting pain in video through facial action units," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 41, no. 3, pp. 664–674, 2011.

[22] N. Rathee and D. Ganotra, "A novel approach for pain intensity detection based on facial feature deformations," *Journal of Visual Communication and Image Representation*, vol. 33, pp. 247–254, 2015.

[23] H. C. Bhakta and C. A. Marco, "Pain management: association with patient satisfaction among emergency department patients," *The Journal of emergency medicine*, vol. 46, no. 4, pp. 456–464, 2014.