Improving Automatic Pain Level Recognition using Pain Site as an Auxiliary Task

1st Hui-Ting Hong
Department of Electrical Engineering,
National Tsing Hua University
MOST Joint Research Center
for AI Technology and All Vista Healthcare
Taiwan

2nd Jeng-Lin Li
Department of Electrical Engineering,
National Tsing Hua University
MOST Joint Research Center
for AI Technology and All Vista Healthcare
Taiwan

3rd Chun-Min Chang
Department of Electrical Engineering,
National Tsing Hua University
MOST Joint Research Center
for AI Technology and All Vista Healthcare
Taiwan

4th Chi-Chun Lee
Department of Electrical Engineering,
National Tsing Hua University
MOST Joint Research Center
for AI Technology and All Vista Healthcare
Taiwan

***Abstract*—Pain is an unpleasant sensory and distressing feeling usually induced by physical damages, and the intensity is further modulated by the experienced pain site. Objective assessment of pain is critical in a variety of clinical practices, however, the status quo in medical practices is based solely on self-report. Recent advancements have been observed in automatic assessment of pain using audio-video recordings, but most do not consider the complex clinical dependency between pain level and pain site. In this study, we propose a Task Specific Encoder with Soft Layer Ordering structure (TSEN-SLO) that utilizes a learnable tensor to flexibly share information between pain level and pain site while still keeping the representations of each task in their self-encoding layers to improve pain level recognition. Our network learns from both face and voice data and achieves accuracy of 70% and 48.1% in a binary and ternary self-report pain level classification in a challenging in-the-wild setting. The approach improves a relative of 6.5% and 9.1% compare to previous work on the same dataset. Further analysis also demonstrates the variation in the self-reported pain level as observed in the facial and acoustic features for different pain sites, which points toward a potential relationship between the neural-mechanism behind internal pain sensation and its effect on expressive facial/vocal behaviors.**

***Index Terms*—Behavioral Signal Processing (BSP), multi-task learning, triage, pain level, pain site**

I. Introduction

Pain is an extremely prevalent yet complicated symptom [1]. Being an internal yet clinically-relevant sensation, research in the perception of pain [2] and its objective assessment [3], [4] has long been an important research direction. In most cases, the level of pain felt depends not only on the amount of bodily damage but also on one's previous experiences [2]. Pain site, on the other hand, has also been studied to understand the impact of the physicality in pain. In similar pathological processes, different pain locations may lead to different pain experiences. The sensation of pain could be induced by tissues (somatic pain), viscera (visceral pain), nervous systems (neuropathic pain) or even the over-sensitization of the peripheral system (maladaptive pain). These variabilities in pain presents their differences not only in the psychophysics of the sensation but also in neurobiological mechanisms [5], and the relation of pain site with different neural responses have also been extensively studied [6], [7].

Being an important clinical marker, obtaining reliable pain level assessment has played a critical role in the diagnosis and evaluation process across medical applications. Several engineering works have investigated an automated approach to assess the pain level to mitigate issues of the current clinical use of self-report pain scales. For example, Kaltwang et al. quantify patient's facial expressions in order to differentiate between pain and no pain [8], Ashraf et al. utilize the active appearance model to recognize frame-level pain [9], and other researchers also detect pain by modeling body gestures [10], [11]. Recently, a series of studies also show that speech modality also possesses substantial information for objective assessment of pain [12], [13].

Previous research has shown that the location of pain may further enhance the pain sensation for patients with both chronic and acute pain [14]. Nevertheless, none of the automated recognition frameworks has considered the complex relationship between pain site and pain level although both information is often collected together for clinical decisions. In our work, we propose a multi-task network of a soft layer ordering structure combined with task specific encoder (TSEN-SLO), where the latent relationship between pain level and pain site is captured through a learnable tensor while each task still retains specific information in the self-encoding branch. We evaluate our framework on a large-scale audio-video corpus collected during real in-hospital emergency room

978-1-7281-3891-6/19/$31.00 ©2019 IEEE

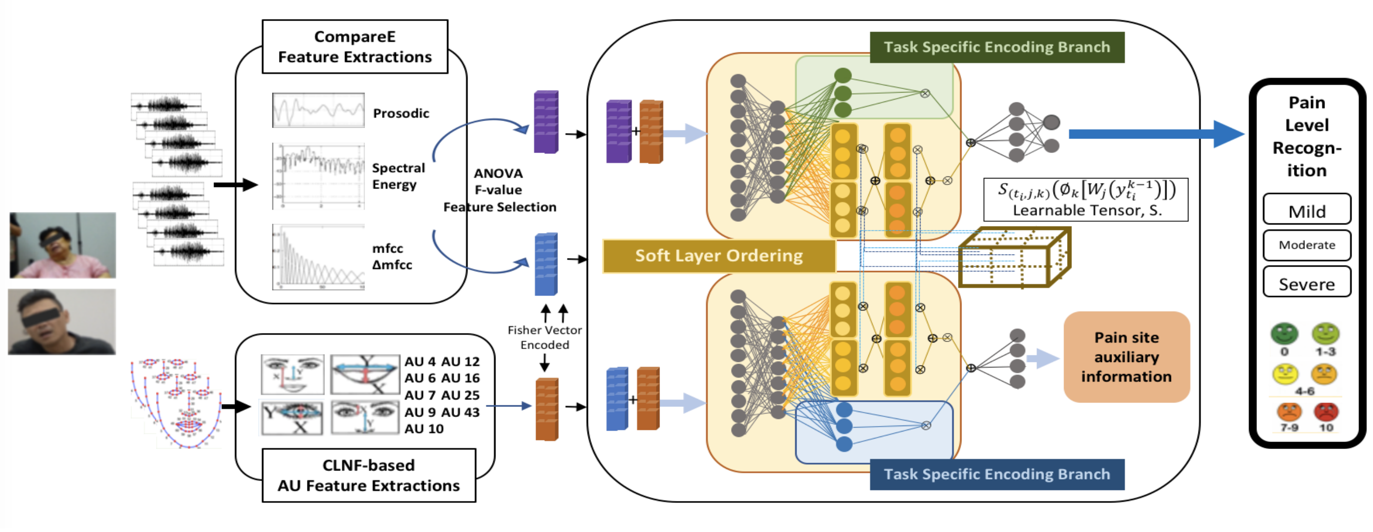

Fig. 1. The complete architecture of our proposed Task Specific Encoder with Soft Layer Ordering (TSEN-SLO) in automatic pain level recognition: extracted acoustic and facial low-level descriptors, session level encoding using Gaussian mixture model based Fisher Vector (GMM FV), training the soft layer ordering network with pain site as an auxiliary task and pain level as the main task.

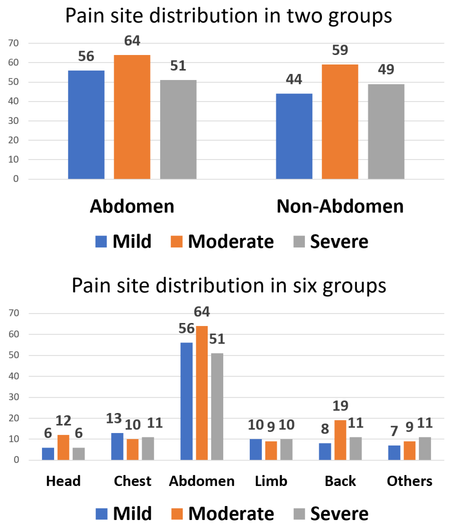

Fig. 2. The data distribution of pain level with different pain sites

triage sessions [15]. Our proposed TSEN-SLO achieves 70% in binary pain level classification (severe vs. mild), 48.1% in a ternary pain level classification (mild, moderate, severe) using face and voice features. We further conduct statistical testing reveals that 1) facial expressions demonstrate differences between severe and mild pain level especially when patients suffer from somatic pain, 2) spectral related features, especially the RASTA style auditory spectrum on the higher mel-frequency bands, display higher mean value in severe pain, and 3) voicing related acoustic features (jitter, shimmer) shows a deduction in values from mild to severe pain for patients suffered from headache.

II. Research Methodology

A. The Triage Audio-Video Pain Database

We utilize the triage pain level database in this work [15]. It was collected at the Department of Emergency at Chang Gung Memorial Hospital, which included audio-video recordings (with manual utterance segmentation), physiological (heart rate, systolic and diastolic blood pressure) data, and other clinically-related outcomes of on-boarding emergency patients during real triage session. Triage nurses recorded each patient's location of pain (pain site) and NRS of pain level, which is the current clinical practice in quantifying pain level on a 10-point self-report scale [16].

TABLE I

The list of extracted facial features modified from [15] are shown below. The number in the parenthesis corresponds to the number of descriptors (25 facial features per frame in total).

Facial action unit inspired descriptors	Descriptions
Frown (1)	The distance between brow
Eyebrow var. (4)	The distance between brow/eyebrow tail and eye
Squint eyes (4)	The distance between upper/lower eyelid and center of eyes height
Eyes opening (2)	The distance between upper/lower eyelid
Eyes var. (1)	The distance between inner eyes
Mouth var. (2)	The height/width of mouth
Cheek var. (2)	The distance between eye and corners of the lips
Upper/Lower lip var. (2)	The distance between upper/lower lip and eyes center
Mouth aspect ratio (1)	The distance of mouth opening
Eyes aspect ratio (2)	The distance of eyes opening
Philtrum var. (1)	The length of noise to philtrum
Nasolabial var. (1)	The width of nasolabial
Head ratation (2)	The degree of rotation between y-axis and z-axis

In this work, we use a total of 323 samples (184 unique patients) in the database. It is the same setting as the most recent automatic pain level recognition work on this corpus [13]. The pain level score is categorized into three commonly-used pain levels, which are *mild*: 0-3, *moderate*: 4-6 and *severe*: 7-10. The pain site is originally categorized into six categories (head, chest, abdomen, back, limb, and others), and considering the different neural-mechanism behind internal pain sensation we further group pain site into two categories: abdomen pain and non-abdomen pain. A plot on the distribution between pain site and pain level is shown in Fig. 2.

B. Pain Level Recognition Framework

Fig. 1. depicts our complete proposed framework of soft layer ordering with task-specific network (TSEN-SLO) for automatic pain level recognition using audio and video features. It models vocal and facial behaviors and further models the latent relationship between pain level (main task) and pain site (auxiliary task). In the following, we will detail the acoustic and facial feature extraction, session level feature encoding, and finally our proposed recognition architecture.

1) Acoustic Features: The acoustic feature set used here is originally developed for interspeech computational paralinguistics challenge [18]. We extract a variety of acoustic low-level (frame-level) descriptors (LLD) with extensive statistical functions to compute an utterance-level feature vector using the openSMILE toolkit [17]. Specifically, it contains 6373 acoustic features per utterance covering different aspect in speech: prosodic, spectral, cepstral and voice quality.

2) Facial Features: Facial action unit features have been demonstrated to be related to pain [19], e.g., AU4, 6, 7, 9, 10, 12, 16, 25, 43. In this work, we follow from previous work in designing 25 facial action unit-inspired low-level descriptors per frame (a modification from [15]). These features characterize eyes, mouth, eyebrows, and nose movements based on the automatically tracked 68 facial landmarks extracted using methods of constrained local neural fields (CLNF) [20]. The details of the 25 features are listed in TABLE I.

3) Session-level Behavior Encoding: Each triage session consists of different lengths of sequences in the extracted audio and video features. We further perform session-level encoding to obtain a fixed-dimensional vectorized representation for each patient at each triage. The session-level encoding that is utilized is based on a method of Gaussian Mixture Model Fisher Vector encoding (GMM-FV) [21]. This particular method has shown its modeling power for tasks of speech paralinguistics recognition [22]. A brief description is below.

Assume our input, X (the behavior features we have extracted), can be modeled using a probability density function p with parameters λ. Here we choose p as a Gaussian Mixture Model (GMM), which denotes as $p(x) = \sum_{i=1}^{K} w_i * p_i(x)$, and $\lambda = \{w_i, \mu_i, \sigma_i, i = 1...K\}$ where K is the number of mixtures, w_i, μ_i and σ_i are the mixture weight (priors), mean vector and diagonal covariances of Gaussian i, respectively. This leads to the Fisher vector encoding by computing the first and second order differentiation between a feature input to each center within the GMM:

$$\Phi_k^{(1)} = \frac{1}{N\sqrt{w_k}} \sum_{d=1}^{N} \alpha_d(k)(\frac{x_d - \mu_k}{\sigma_k}) \tag{1}$$

$$\Phi_k^{(2)} = \frac{1}{N\sqrt{2w_k}} \sum_{d=1}^{N} \alpha_d(k)(\frac{(x_d - \mu_k)^2}{\sigma_k^2} - 1) \tag{2}$$

where N is the feature dimension, x_d stands for the d^{th} dimension of input X and $\alpha_d(k)$ defined as :

$$\alpha_d(k) = \frac{w_i p_i(x_d)}{\sum_{j=1}^{K} w_j * p_j(x_d)} \tag{3}$$

The encoded fisher vector (FV), ϕ, can then be obtained by stacking the Φ_k: $\phi = [\Phi_1^{(1)}, \Phi_1^{(2)}, ..., \Phi_k^{(1)}, \Phi_k^{(2)}]$. This serves as our input to the TSEN-SLO network.

4) Task Specific Encoder with Soft Layer Ordering (TSEN-SLO): We propose a TSEN-SLO network architecture that models the latent relationship between pain level and pain site. The structure is mainly a soft layer ordering multitask structure in learning to assemble a set of shared layers with weights modified using learnable tensors for different tasks [23]. We apply a similar two-branch network for our tasks, i.e., pain level and pain site, by designing a core network with two learned affine layers, $W_1, W_2 : R^m \rightarrow R^m$, that are shared across the two tasks. Based on the soft layer ordering, we define the equations for the two tasks as follows:

$$y_1^k = \sum_{j=1}^{D} S_{(1,j,k)}(\phi[W_j(y_1^{k-1})]) \tag{4}$$

$$y_2^k = \sum_{j=1}^{D} S_{(2,j,k)}(\phi[W_j(y_2^{k-1})]) \tag{5}$$

where $\phi = ReLU$, D equals to two in this experiment, and each element of $S_{(t_i,j,k)}$ is drawn from a learnable tensor, S, to derive a scalar at depth k that modifies layer W_j for task t_i:

$$\sum_{j=1}^{D} S_{(t_i,j,k)} = 1 \quad \forall (t_i, k), \quad t_i = 1, 2 \tag{6}$$

We further introduce a novel use of an additional task-specific branch to help the model adapt the learned shared information to each task more precisely. Depth of k therefore extend from range [0,1] to [0,2]. We can then rewrite the original equation (4) and (5) as follows:

$$y_{t_i} = S_{(t_i,2,2)} \cdot \phi(F_{t_i}(y_{t_i}^0))$$
$$+ S_{(t_i,1,2)} \cdot (\sum_{j=1}^{2} S_{(t_i,j,0)}(\phi(y_{t_i}^0)) + \sum_{j=1}^{2} S_{(t_i,j,1)}(\phi[W_j(y_{t_i}^0)]))$$

where $y_{t_i}^0$ is defined as output before the core shareable portion of the network, and F_{t_i} is the task-specific affine layers: $R^m \rightarrow R^m$. The tensor S is learned jointly during backpropagation. The network essentially learns a scalar-modified weights drawn from a depth-dependent tensor dictating how these two shared task-layers are assembled. Further, with the use of additional task-specific encoding layers for each task separately, the TSEN-SLO provides flexibility in extracting both shared yet task-specific representation. In summary, TSEN-SLO uses a learnable tensor that is jointly optimized within the shared task layers that combines with task-specific branch to further improve the adaptability for enhanced modeling power.

TABLE II

It summarizes the detailed results of the baseline and the multitask recognition result, where PL indicates Pain-Level and PS indicates Pain-Site. We include VAE, STL, Hard-S, Soft-S, SLO approaches as our comparison methods. The number with bold means the best result in our experiment.

	VAE	*STL*			*Hard-S*		*Soft-S*		*SLO*		*TSEN-SLO*	
	Vocal	*Vocal*	*Facial*	*Vocal+Facial*	*Vocal+Facial*							
PL: 2-Class					PS = 2	PS = 6	PS = 2	PS = 6	PS = 2	PS = 6	PS = 2	PS = 6
Mild	66.0	63.0	58.5	64.0	69.0	65.0	68.0	68.0	72.0	69.0	76.0	73.0
Severe	61.0	63.0	58.0	63.0	63.0	62.0	63.0	60.0	58.0	62.0	64.0	61.0
UAR	63.5	63.0	58.5	63.5	66.0	63.5	65.5	64.0	65.0	65.5	**70.0**	68.0
PL: 3-Class												
Mild	43.7	42.0	43.0	46.0	45.0	46.0	49.0	51.0	46.0	46.0	50.0	52.0
Moderate	35.4	42.2	41.5	43.1	45.5	46.3	41.4	42.3	48.0	48.8	49.5	45.5
Severe	38.1	36.0	39.0	43.0	39.0	40.0	41.0	43.0	48.0	41.0	45.0	41.0
UAR	39.1	40.0	41.2	44.0	43.1	44.1	43.8	45.4	45.7	45.3	**48.1**	46.2

III. Experimental Setup and Results

A. Experimental Setup

In this work, we present recognition results both in pain level (binary and ternary) by using the auxiliary task of pain site (binary and six classes). All evaluation is done via leave-one-patient-out cross-validation using the unweighted average recall (UAR) as the performance metric. Our TSEN-SLO structure is composed of four hidden layers and three-dimensional tensor variables. All variables are initialized with uniform distribution, and the ReLU activation function is applied. The hyperparameters list: 20 for batch size, 0.001 for learning rate, Adam optimizer, and 20 epochs. We perform univariate ANOVA feature selection for each modality of each task (pain level and pain site) separately before performing GMM FV to reduce the dimension of the encoded session level representations. Both encoded acoustic and video features are concatenated as input to the proposed TSEN-SLO. We further compare our proposed method with the following models:

- **VAE [13]**: Variational learning with Maximum-Mean Discrepancy criterion training on the acoustic LLDs.
- **STL**: Single task learning for pain level recognition.
- **Hard-S**: Hard parameter sharing, hidden layers for the two tasks are shared, and branch out at the output layer for task-specific recognition.
- **Soft-S**: Soft parameter sharing, each task has its parameters in the hidden layers and is learned through joint optimization for both tasks.
- **SLO**: Soft layer ordering, proposed by [23].
- **TSEN-SLO**: Our proposed method.

Note that, the Hard-S and Soft-S are the two standard multi-task learning structures [24].

B. Experimental Results

TABLE II summarizes our pain level recognition results as a function on the different number of pain sites used as an auxiliary task. Our proposed TSEN-SLO achieves the best accuracy of 70% and 48.1% in 2-class and 3-class classification on pain level. i.e. a gain of 6.5%, 4.1% on pain level in 2-class and 3-class, respectively compared to multimodal STL (vocal+facial single task learning). This result also surpasses the most recent work on the same dataset that applies a variational learning approach in the speech modality [13] by 6.5% and 9.1% in 2-class and 3-class recognition results.

There are several notable observations to be made. First, in general, we observe that the fusion of vocal/facial modalities provides improvement in the pain level assessment. Second, we see that even with simple multi-task learning structure, i.e., both Hard-S and Soft-S structures, would improve recognition rates compared to single-task learning indicating there indeed exists complementary latent relationship between pain level and pain site as manifested in the patient's vocal/facial behaviors.

Third, the SLO structure result shows the power in flexibly learn the shared representation between the two tasks to improve main tasks recognition rates compared to Hard-S and Soft-S. Finally, SLO combining with the additional use of task-specific layer (i.e., our proposed TSEN-SLO) adds additional modeling power, which increases 5% and 2.4% compared to SLO on 2-class and 3-class pain level recognition. In summary, our proposed TSEN-SLO flexibly learns the shared information and provides additional modeling power that makes the network capacity to be adaptive to each task simultaneously. In a real-world multi-task problem, where the relationship between the tasks (e.g., pain level and pain site) can be extremely complicated, learning how the common representation layers should be shared in different depths for different tasks with an extra task-specific branch to handle specificity in each task is beneficial in the overall recognition tasks.

IV. Analysis of Behavior Differences between Pain Level across Pain Site

In this section, we further analyze the differences of each behavior modality across pain levels under each pain site condition. We first compute five statistic functions (mean, max, min, standard deviation) on each patient's utterance. Then, a two-sided Student's t-test is performed between pain level groups (severe vs. mild) with 0.05 significant level as cutoff under each condition of pain site. The statistical testing result is summarized in TABLE III.

A. Analysis on Facial Expressions

The significant differences between severe and mild pain level are observed in abdomen, limb, and others pain sites for

TABLE III

A table summarize the two-sided Student's t-tests. The listed features are significantly different between groups of pain level : severe pain versus mild pain. All of the listed descriptors below obtain p-value less than .05

FACIAL EXPRESSIONS		ACOUSTIC EXPRESSIONS	
Severe > Mild	**Severe < Mild**	**Severe > Mild**	**Severe < Mild**
Abdomen Pain Site		*Head Pain Site*	
Head Rotation 2nd-(min)	Squint eyes 2nd-(min)	RAS.audSpec-bands 24th-(std,max)	jitterLocal-(mean)
Limb Pain Site		RAS.audSpec-bands 23th-(max)	jitterDDP-(mean)
Upper/Lower lip Var. 1st-(std)	–	RAS.audSpec-bands 25th-(max)	shimmerLocal-(mean)
Philtrum Var.-(std)	–	–	RAS.audSpec-bands 6th,7th,12th,13th-(mean)
Others Pain Site		–	RAS.audSpec-bands 12th,13th-(std)
Head Rotation 2nd-(mean)	Eyebrow Var. 1st,2nd,3rd-(std)	–	mfcc 5th-(std)
–	Mouth Var. 2nd-(std)	*Abdomen Pain Site*	
–	Cheek Var. 1st,2nd-(std)	voicingFinalUnclipped-(std)	mfcc 11th-(max)
–	Upper/Lower lip Var. 1st-(std)	*Limb Pain Site*	
–	Nasolabial Var.-(std)	RAS.audSpec-bands 10th-(std)	–
–	Eyebrow Var. 1st-(max)	*Back Pain Site*	
–	Philtrum Var.-(max)	–	voicingFinalUnclipped-(min)
–	Nasolabial Var.-(max)		

Var. stand for variation; detail explained in TABLE I

jitter/shimmerLocal stands for the average absolute difference between two consecutive periods

(the former measures the acoustic periods and the latter on the amplitudes of the fundamental frequency.);

jitterDDP stands for the average absolute difference between jitter cycles; RAS.audSpec stands for RASTA style auditory spectrum;

VoicingFinalUnclipped stands for the voicing probability of the final fundamental frequency candidate with no zero-clipping when falls below a voicing threshold.

facial expressions. Among six different pain sites, patients reporting on the neck, right/left shoulder and lower quadrant are categorized as 'others' pain site; this type of pain is clinically considered as the somatic pain [25] which highly relates to muscle and joint problems [26]. Our result reveals a significant difference for seven kinds of facial expression measures in the 'others' pain site condition; the seven types of video features include *head rotation, eyebrow variation, mouth variation, cheek variation, upper/lower lip variation, nasolabial variation* and *philtrum variation*. Since the sensation of pain could predominantly affect the muscles and their association has long been developed [27], it is intuitively pleasing to see most muscular expression on face showing differences between different levels of pain reported, especially under the 'others' pain site condition. Abdomen pain, in contrast, is more likely to be caused by visceral organ pathology [28], where only two facial expressions (*head rotation, squint eyes*) display the significant difference between pain level groups.

B. Analysis on Vocal Characteristics

TABLE III shows statistically significant difference for five acoustic measures in the 'head' pain site condition, which are *RASTA-style auditory spectrum, jitterLocal, jitterDDP, shimmerLocal and mfcc*. RASTA-style auditory spectrum bands 1-26 indicates the raw values per band of RASTA filtered auditory band levels [1], the intuitive finding shows that auditory spectrum applied to higher frequency (RASTA-style auditory spectrum bands 23th,24th,25th) display higher mean in severe pain compare to mild pain. On the other hand, lower frequency (RASTA-style auditory spectrum bands 6th,7th,12th,13th) shows a lower mean in severe pain.

Another surprising observation is found that voicing related acoustic features (*jitterLocal, jitterDDP, shimmerLocal*) demonstrate a reduction in value from mild to severe pain (i.e. pathological-related jitter and shimmer values are less in severe pain than in mild pain). In other words, when patients suffer from symptoms of headache, the variation on the fundamental frequency and energy for phonation (both results from the physical muscular adjustments of the larynx) are negatively correlated to the pain intensity; while none of any other pain site (except head pain site) shows mean differences of these voicing related parameters. This intriguing result implies that patients with nociceptive pain that is transmitted from different pathway may likely to cause the patient's acoustic characteristics to alter in a non-intuitive manner.

C. Discussions

Both vocal and facial expressions provide critical windows in reflecting the internal sensation of pain for patients. In this study, we demonstrate that measures of facial expressions show a clear distinction between painful and non-painful scenarios, and so as in acoustic manifestation. While pain site and pain level both serve as important clinical variables in medical practices, little effort has been made to systematically understand and model the dependency between these two factors and understand how they impact behavior manifestations. In this work, we observe that when patients suffer from somatic pain, facial expressions of *head rotation, eyebrow variation, mouth variation, cheek variation, upper/lower lip variation, nasolabial variation* and *philtrum variation* display significant

[1] RASTA style bandpass filter applied to Mel-frequency bands (0-8kHz)

differences between severe and mild pain. Spectral related acoustic features show differences when patients experiencing pain sites of head, abdomen, and limb, while voicing related features demonstrate differences between pain level when reporting pain in the head. In other words, as patients suffer from different regions of pain, different variability of facial and acoustic behaviors could imply a different level of pain intensity. This particular subtly dependency between pain site and pain levels can also be leveraged in improving automatic pain level assessment as demonstrated in using our proposed TSEN-SLO framework.

V. Conclusions and Future works

In many medical applications, knowing pain site and pain level are both critical to the evaluation of treatment. This work presents an modeling on the latent relationship between the two factors in a real clinical setting of emergency triage database. By proposing a task-specific encoding layer combined with soft layer ordering (TSEN-SLO) structure, we achieve 70% and 48.1% on automatic 2-class, 3-class pain-level recognition. We demonstrate that facial expressions show differences between severe vs. mild pain especially when patients are suffering from somatic pain, and surprisingly we observe that a counter-intuitive result where patients display the higher value of pathology-related voice quality measures (i.e., jitter and shimmer) for mild pain condition during headache. To the best of our knowledge, this is one of the first work that leverages and studies the relationship between pain level and pain site from the multimodal behavior perspective. In addition, we will continue to advance the technical aspect of our network architecture with explicit loss constraint embedded into the shared soft layer and task-specific encoder structure to improve further the recognition rate, and further analyze the soft layer ordering weights to bring additional insights to the different pain-related neurobiological mechanism and its relationship to the vocal and facial expressions.

References

[1] Ronald Melzack and Kenneth L Casey, "Sensory, motivational and central control determinants of pain: a new conceptual model," *The skin senses*, vol. 1, 1968.

[2] Ronald Melzack, "The perception of pain," *Scientific American*, vol. 204, no. 2, pp. 41–49, 1961.

[3] Harald Breivik, PC Borchgrevink, SM Allen, LA Rosseland, L Romundstad, EK Breivik Hals, G Kvarstein, and A Stubhaug, "Assessment of pain," *British journal of anaesthesia*, vol. 101, no. 1, pp. 17–24, 2008.

[4] Tim McCormick and Simon Law, "Assessment of acute and chronic pain," *Anaesthesia & Intensive Care Medicine*, vol. 17, no. 9, pp. 421–424, 2016.

[5] Fernando Cervero, "Visceral versus somatic pain: similarities and differences," *Digestive Diseases*, vol. 27, no. Suppl. 1, pp. 3–10, 2009.

[6] Kischner S and McMyne RC, "Dermatomes anatomy," 2015.

[7] Patrick W Mantyh, "The neurobiology of skeletal pain," *European journal of Neuroscience*, vol. 39, no. 3, pp. 508–519, 2014.

[8] Sebastian Kaltwang, Ognjen Rudovic, and Maja Pantic, "Continuous pain intensity estimation from facial expressions," in *International Symposium on Visual Computing*. Springer, 2012, pp. 368–377.

[9] Ahmed Bilal Ashraf, Simon Lucey, Jeffrey F Cohn, Tsuhan Chen, Zara Ambadar, Kenneth M Prkachin, and Patricia E Solomon, "The painful face–pain expression recognition using active appearance models," *Image and vision computing*, vol. 27, no. 12, pp. 1788–1796, 2009.

[10] Temitayo A Olugbade, Aneesha Singh, Nadia Bianchi-Berthouze, Nicolai Marquardt, Min SH Aung, and Amanda C De C Williams, "How can affect be detected and represented in technological support for physical rehabilitation?," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 26, no. 1, pp. 1, 2019.

[11] Jesús Joel Rivas, Felipe Orihuela-Espina, Lorena Palafox, Nadia Berthouze, María del Carmen Lara, Jorge Hernández-Franco, and Enrique Sucar, "Unobtrusive inference of affective states in virtual rehabilitation from upper limb motions: A feasibility study," *IEEE Transactions on Affective Computing*, 2018.

[12] Fu-Sheng Tsai, Yi-Ming Weng, Chip-Jin Ng, and Chi-Chun Lee, "Embedding stacked bottleneck vocal features in a lstm architecture for automatic pain level classification during emergency triage," in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 2017, pp. 313–318.

[13] Jeng-Lin Li, Yi-Ming Weng, Chip-Jin Ng, and Chi-Chun Lee, "Learning conditional acoustic latent representation with gender and age attributes for automatic pain level recognition," *Proc. Interspeech 2018*, pp. 3438–3442, 2018.

[14] Andrew Rice, *Clinical pain management*, Hodder Arnold, London, 2008.

[15] Fu-Sheng Tsai, Ya-Ling Hsu, Wei-Chen Chen, Yi-Ming Weng, Chip-Jin Ng, and Chi-Chun Lee, "Toward development and evaluation of pain level-rating scale for emergency triage based on vocal characteristics and facial expressions.," in *INTERSPEECH*, 2016, pp. 92–96.

[16] Kerstin Eriksson, Lotta Wikström, Kristofer Årestedt, Bengt Fridlund, and Anders Broström, "Numeric rating scale: patients' perceptions of its use in postoperative pain assessments," *Applied nursing research*, vol. 27, no. 1, pp. 41–46, 2014.

[17] Florian Eyben, Martin Wöllmer, and Björn Schuller, "Opensmile: the munich versatile and fast open-source audio feature extractor," in *Proceedings of the 18th ACM international conference on Multimedia*. ACM, 2010, pp. 1459–1462.

[18] Björn Schuller, Stefan Steidl, Anton Batliner, Alessandro Vinciarelli, Klaus Scherer, Fabien Ringeval, Mohamed Chetouani, Felix Weninger, Florian Eyben, Erik Marchi, et al., "The interspeech 2013 computational paralinguistics challenge: social signals, conflict, emotion, autism," in *Proceedings INTERSPEECH 2013, 14th Annual Conference of the International Speech Communication Association, Lyon, France*, 2013.

[19] Patrick Lucey, Jeffrey F Cohn, Iain Matthews, Simon Lucey, Sridha Sridharan, Jessica Howlett, and Kenneth M Prkachin, "Automatically detecting pain in video through facial action units," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 41, no. 3, pp. 664–674, 2011.

[20] Tadas Baltrusaitis, Peter Robinson, and Louis-Philippe 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.

[21] Florent Perronnin, Jorge Sánchez, and Thomas Mensink, "Improving the fisher kernel for large-scale image classification," in *European conference on computer vision*. Springer, 2010, pp. 143–156.

[22] Heysem Kaya, Alexey A Karpov, and Albert Ali Salah, "Fisher vectors with cascaded normalization for paralinguistic analysis," in *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.

[23] Elliot Meyerson and Risto Miikkulainen, "Beyond shared hierarchies: Deep multitask learning through soft layer ordering," *arXiv preprint arXiv:1711.00108*, 2017.

[24] Sebastian Ruder, "An overview of multi-task learning in deep neural networks," *arXiv preprint arXiv:1706.05098*, 2017.

[25] R. Sherman, "Abdominal pain," 1990.

[26] K. Sluka, "Stimulation of deep somatic tissue with capsaicin produces long-lasting mechanical allodynia and heat hypoalgesia that depends on early activation of the camp pathway," *Journal of Neuroscience*, vol. 22, no. 13, pp. 5687–5693, 2002.

[27] D. Taverner, "Muscle spasm as a cause of somatic pain," *Annals of the rheumatic diseases*, vol. 13, no. 4, p. 331, 1954.

[28] S. Sikandar and A. H. Dickenson, "Visceral pain–the ins and outs, the ups and downs," *Current opinion in supportive and palliative care*, vol. 6, no. 1, p. 17, 2012.

[29] F. Weninger, F. Eyben, B. W. Schuller, M. Mortillaro, and K. R. Scherer, "On the acoustics of emotion in audio: what speech, music, and sound have in common," *Frontiers in psychology*, vol. 4, p. 292, 2013.