Speech Prosody 2016
31 May - 3 Jun 2106, Boston, USA

# Speech Prosody as a Biosignal for Physical Pain Detection

*Yaniv Oshrat*[1], *Ayala Bloch*[2], *Anat Lerner*[1], *Azaria Cohen*[1], *Mireille Avigal*[1], *Gabi Zeilig*[3,4]

[1]Department of Mathematics and Computer Science, the Open University of Israel
[2]The National Institute for the Rehabilitation of the Brain Injured, Israel
[3]Department of Neurological Rehabilitation, the Chaim Sheba Medical Center Tel-Hashomer
[4]Sackler Faculty of Medicine, Tel-Aviv University

oshblo@zahav.net.il, ayalabl@shikumil.org.il, anat@cs.openu.ac.il,
azaria.cohen@gmail.com, miray@openu.ac.il, gabi.zeilig@sheba.health.gov.il

## Abstract

Obtaining an objective assessment of pain is an important challenge for clinicians. The purpose of this study is to examine the connections between subjective reports of pain and measureable biosignals of human speech prosody, as a step towards coping with this challenge.

Patients reporting pain were voice-recorded to attain reports on different levels of pain. Recording was done in the patients' natural environment at the medical center. Features were extracted from the voice-recordings, including features that were exclusively developed for this study. A machine-learning based classification process was performed in order to distinguish between samples with "no significant pain" and with "significant pain" reported. This classification process distinguished well between the two categories. Moreover, features developed during this study improved classification results in comparison to classification based solely on known-features.

Results indicate that there is evidence of a connection between measureable biosignal parameters of speech and the simultaneous self-reported pain level. This finding might be useful for developing future methods to more objective assessment of pain.

**Index Terms**: Signal processing, machine learning, statistical classifiers, speech prosody in pain.

## 1. Introduction

Pain is an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage [1].

Researchers have shown connections between pain and measureable physical parameters. Loggia, Juneau and Bushnell [2] studied heart rate and skin conductance in response to pain inflicted as heat stimuli, combined with levels of pain rated by the subjects. They found that skin conductance and heart rate significantly increase during pain. This result confirms a wide range of previous studies that have observed this relationship with either heart rate, skin conductance, or both, using many types of pain.

In addition to the above, several other studies showed relations between heart rate and systolic blood pressure, and pain [3], [4]. In all cases, stimuli or exercise caused an increase in the physiologic parameters measured.

Voice might also be among the physiological parameters connected to pain. Studies have shown that the human voice is affected by physiological factors. Orlikoff and Baker [5] found that the cardiovascular system has a consistent affect on vocal fundamental frequency (F0). Orlikoff [6] showed that vowel amplitude variation is affected by the cardiac (ECG) cycle.

Human voice and prosody contains vast information, and many studies have shown that extensive knowledge about a speaker's physiology can be extracted from the voice [7], [8]. In recent years numerous attempts have been made to extract similar information about a speaker's emotional state as well [9], [10], [11].

Recognizing that pain affects physiological parameters (e.g. heart rate [12]) of the human body, and that physiological characteristics of the body (e.g. heart rate [5]) influence the voice; it should be interesting to explore whether there is a connection between the subjective experience of pain and a person's voice. Since speech is the fastest and most natural means of human communication, this possible connection may have practical implications, for instance assistance in medical diagnosis, in assessment of the patient's complaints or in acquisition of a pain report when verbal communication with the patient is limited. Currently, clinicians base their pain assessment exclusively on the patient's subjective pain report. This approach might evolve in the future into a medical aide that will provide clinicians with more objective diagnostic tools.

The current study is a preliminary investigation that aims to revealing a connection between subjective self-report of pain and parameters of speech prosody at the time of the report. For this purpose we first considered features that are commonly used in emotional speech research. We then developed some new exclusive features for pain situations.

We are aware of the limited quantity of speech (400 short samples), and we regard this study as a Proof of Concept (POC) study and further research is due.

We proceed as follows: in Section 2, we describe the research methods used; In Section 3 we pose the research questions; In Section 4 we describe the exclusively developed features and their motivation; In Section 5 we show and analyze the results; and finally we conclude in Section 6.

 doi: 10.21437/SpeechProsody.2016-86

# 2. Research Methods

The study was conducted by orally interviewing patients in the Inpatient Department of Neurological Rehabilitation and the Outpatient Mobility Rehabilitation Clinic at the Chaim Sheba Medical Center, Tel-Hashomer, Israel.

## 2.1. Participants

Participants included 27 adults (20 men, 7 women, age range: 23-65 years, mean age: 44 years). All of the participants had spinal cord and/or brain injuries, and reported injury-related pain. No pain, physical pressure or any intrusive action was inflicted upon participants as part of the study.

### *2.1.1. Natural environment principal*

Participants were approached in their natural environment at the medical center, and in their regular day-to-day course of behavior. Daily schedule, including sleep, meals, physical and psychological therapies and other elements of rehabilitation program, was maintained intact. Participants taking medicines (pain killers or other kind of drugs) continued to take them regularly during the study. [1]

## 2.2. Data Collection

The data collection took six months. Each session included a short voice-recorded interview.

The interview took place in a closed location that enabled quiet and privacy in the participant's surrounding. The same person conducted all the interviews with all the participants.

In the interview, participants were asked to state their names and a string of 9 digits that they know by heart. Then they were asked to evaluate their current level of pain on a standard numeric scale from 1 (no pain) to 10 (worst pain ever experienced). This pain scale is the standard scale used both in research and in clinical practice in the Sheba Medical Center. In some cases participants stated a two-level pain degree (e.g. 5-6), in which case the mid-level degree (i.e., 5.5) was annotated.

The interviews were recorded using the ZOOM Handy Recorder Type H4n (Japan), in stereophonic recording, defining a sampling rate of 96 kHz.

After the recording phase, interviews were listened to and checked to verify recording quality and lack of significant surrounding noises. The total number of interviews that were found to comply with the study requirements was 97.

## 2.3. Sample processing

Each interview was cut into short utterances of either string of digits or words. The mean length of a sample was 0.93 seconds (min. 0.23sec, max. 2.00sec). These characteristics were chosen in order to take similar samples from each participant in all the interviews, and in order to ensure the respone was a natural answer to a question and not a response given in a reciting mode: the string of digits and the name do not change and the participant knows them by heart, without any external assistance.

Cutting the recorded interviews into voice samples was done manually using Audacity 2.0.5 software for voice editing. All the interviews of a specific participant were cut the same way into samples that included the same string of digits or words. From a single interview, 3 to 6 samples were derived in this manner, depending on the participant's speed of speech, pronunciation and choice of words. After removing the damaged samples (surrounding noises, speech mistakes, etc.) 400 samples remained from 97 interviews of the 27 participants.

## 2.4. Feature extraction

Audio features from the recorded samples were extracted using openSMILE (the Munich open Speech and Music Interpretation by Large Space Extraction toolkit), version 1.0.1 [13]. The openSMILE 'emobase 2010' reference set of features, which is based on the INTERSPEECH 2010 Paralinguistic Challenge feature set, was used. This set is very comprehensive and contains 1582 features, based on 34 low-level descriptors (LLD), corresponding delta coefficients and functionals, and is recommended by the developers as a reference feature set [14].

## 2.5. Machine learning analysis

Analysis of the data was done using WEKA (Waikato Environment for Knowledge Analysis) suite, version 3.6.10 [15]. We used the attribute selection feature of WEKA to choose the most effective attributes from the long list of features presented by 'emobase 2010' reference set, using the Correlation-based Feature Selection (CFS) [16], implemented in WEKA as CfsSubsetEval. We used the classifier SMO (based on the supervised learner SVM), with cross validation and five folds.

# 3. Research questions

The original research question Q1 was: Can we classify the level of pain based on acoustic features? We found no significant classification. This finding was consistent with other attempts to differentiate between several categories of pain using physiological parameters [12].

We therefore modified the research question to Q2: Can we distinguish "significant pain" from "no significant pain". As pain is a relatively subjective experience [1], and according to the medical staff recommandation, the separation line was chosen to be: "no significant pain" includes all pain levels lower or equal to 2; "significant pain" includes all pain levels higher or equal to 2.5.

Using this separation-line, the data was divided into 271 samples with significant pain and 129 samples with no significant pain.

Reasearch question Q3 related to the exclusively developed features described in Section 4. The question was: can these features further improve the classification?

# 4. Exclusively developed features

We tried to develop more features that are not included in 'Emobase 2010' feature set, and that might comply better for

[1] Permission was obtained from the Sheba Medical Center ethical committee before recruiting participants.

the situation of pain. Two types of features were developed: Sequence Indication and Heat Map Thresholding.

### 4.1. Sequence Indication

Several studies, regarding different kinds of pain, showed that one of the symptoms of pain is making human responses less rapid [17], [18], [19]. This symptom might be expressed by an alteration in the rate of change in some voice parameters, which in turn may modify the length of sequences of relatively closed values of these parameters. Sequence, in this context, is a list of consecutive measurements of given parameters (the parameters will be discussed below) that are close in value to each other, or residing in a close neighborhood. The term "close" is relative to a pre-defined ε-value; two values V1 and V2 will be considered close if and only if $|V1-V2|<\varepsilon$.

We developed two kinds of sequence-length indications for lists of values: α-sequence and β-sequence indications that are typical to audio samples. In α-sequence indication, the sequence starts at the first measurement and ends when one measurement is not close to the previous one. The length of the sequence is the number of measurements in the sequence. In β-sequence indication, the sequence starts at the first measurement and ends when one measure is not close to the first measurement of the sequence. As in the previous case, the length of the sequence is the number of measurements in the sequence. The α-value of a list of measurements is defined as the mean length of all α-sequences in the list. Similarly, the β-value of a list of measurements is defined as the mean length of all β-sequences in the list.

We deployed the sequence indications on the PLP (Perceptual Linear Prediction) analysis that was performed on the voice samples [20]. PLP, like the LPC (Linear Predictive Coding) method on which it is based, is built on a model that approximates the human vocal tract and describes it as a set of filters [21]. Consequently, we presumed that vocal phenomena concerning pain situations might be found in the PLP analysis of speech. Moreover, since we were looking for alterations in the rate of change in voice parameters, we believed that the derivatives would be the right place to investigate.

The PLP analysis produces 13 coefficients for a single time-frame. When PLP is applied to a voice sample (that contains numerous time-frames), 13 vectors are produced: one vector per coefficient. For each of these 13 vectors we computed the first derivative vector (DEL) and the second derivative vector (DDEL). We computed the α-value and the β-value for both vectors. Thus, we computed 52 values for each voice sample (13 PLP coefficient vectors * 2 [DEL + DDEL] * 2 [α-value + β-value] =52). In each computation, the value of ε was set to 1/20 of the range (|max-min|) of the values in the vector (i.e., $\varepsilon=|max-min|*0.05$). Implementation of these measures was made using MATLAB version 7.12.0.365 (R2011a).

### 4.2. Heat map thresholding

The second type of features that was developed is based on Heat Map Thresholding. The motivation for this approach emerged after viewing the heat map image presentations of common feature extraction techniques, such as LPC, MFCC and PLP. In this presentation, the coefficients that were extracted from a voice sample are set in a 2-dimensional matrix: The first dimension is the sequential number of the frame, and the second dimension is the coefficient values of that frame. The matrix is then displayed in a heat map format, using a color scheme that helps to illustrate phenomena in the image. This presentation can be applied to any feature extraction technique that produces a list of coefficients for a single frame. Figure 1 presents a heat map of a voice sample that was produced in the aforementioned way, and colored using the "jet" colormap.

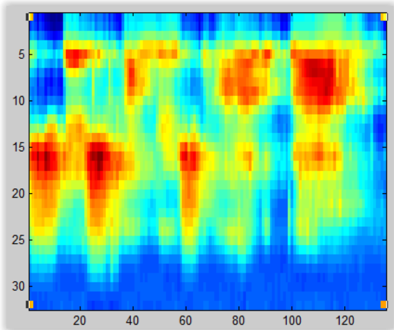


Figure 1: *Voice sample presented as a 2-dimesion matrix of RASTA-PLP coefficients and colored with the "jet" colormap. The X-axis is the sequential number of the frame, the Y-axis presents the coefficients.*

We compared pairs of such images, where each pair was produced by two voice samples that were taken from the same participant. The first sample is with a low pain-level reported and the second is with a high pain-level reported. Visually we noticed differences in the color patterns of the two images, and these differences motivated us to investigate them in a more rigorous manner than merely observing the image.

We produced images using this method for the RelAtive SpecTral PLP (RASTA-PLP) coefficients. RASTA-PLP is an extension of the PLP technique that deals more effectively with problems of real-world recording and communication environment [22]. Similarly to PLP, RASTA-PLP computes the critical-band spectrum and its logarithm, but then it estimates its temporal derivative using regression and spectral values at 5 consecutive time instants, and not just a simple difference [23]. It actually filters the PLP coefficients with a band-pass filter that corresponds to the timing characters of the speech. This operation cancels the convolutional noise that is typical of recordings and communication media. Since we use recorded audio material that was recorded in field conditions, this technique seemed appropriate.

We performed the image analysis in the following manner: The image underwent thresholding with five different threshold values (1/6, 2/6, 3/6, 4/6 and 5/6 in the range of values in the original image). We now had five new images for each voice sample. Each image contained a different number, pattern and size of spots ("spot" is a simply-connected space on the 2-dimensional image). Figure 2 illustrates an original image and its five derivatives.

These two new types of features (α and β sequences indication features, and heat map thresholding features), produce 260 features.

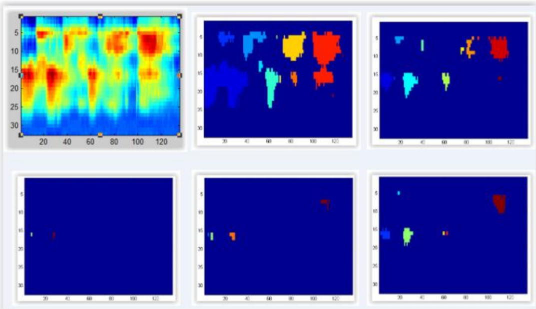

Figure 2: *Original RASTA-PLP 2 dimensional image (upper, left) and its five threshold derivative images (clockwise).*

## 5. Results

We ran two classification experiments of "no significant pain" versus "significant pain". The two experiments varied by the sets of features: The first one included only the features extracted by OpenSMILE, while in the second one we added also the new developed features that we extracted with MatLab. The results are summarized in Table 1.

Table 1. *Classification results.*

| **No. of Instances** | **Patient's pain report** | | **Feature Set** | **Correct classification** | | **Correctly Classified Instances (CCI ratio)** | **Kappa** |
|---|---|---|---|---|---|---|---|
| | Pos | Neg | | True Pos | True Neg | | |
| 400 | 271 | 129 | OpenSMILE | 219 | 76 | 0.7375 | 0.3981 |
| | | | OpenSMILE+ new features | 230 | 79 | 0.7725 | 0.4697 |

The "patient's pain report" column indicates the pain level reported: "Pos" (positive) for "Significant Pain" (i.e. pain level 2.5 to 8); "Neg" (negative) for "No Significant Pain" (i.e. pain level 1 to 2). The Feature Set column indicates which set of features was deployed in the classification. The Correct Classification column indicates the number of "True Positive" and "True Negative" classifications. The Correctly Classified Instances (CCI) ratio presents the percentage of the correctly classified samples out of the total number of samples in the corpus. Kappa is the kappa inter-rater agreement value (a.k.a. "Cohen's kappa", [24]) which serves as a better rater of the classification quality than the CCI [25], [26].

The results in Table 1 show that we can distinguish "significant pain" from "no significant pain" based on acoustic features, even by using the features extracted by OpenSMILE. They also show that the addition of the new features improved the results both with regard to the CCI ratio and the Kappa statistic. Thus the answer to both of the research questions Q2 and Q3 is positive.

Another result of this study concerns the significance of the features: The most effective acoustic features for pain classification from the 'Emobase 2010' feature set were the MFCC (Mel-Frequency Cepstral Coefficients 0-14), the logMelFreqBand (logarithmic power of Mel-frequency bands 0-7) and the lspFreq (8 line spectral pair frequencies computed from 8 LPC coefficients). From the "new features", both sequence indication features and heat map thresholding features were found to be significant.

## 6. Conclusions and Discussion

In this study, we found that speech in pain can be used as a biosignal for the classification of pain experiences, in addition to the patient's report. We classified the voice samples, distinguishing between samples that were uttered when the patient reported having "no significant pain", and samples that were uttered when the patient reported having "significant pain". This classification was done using the recorded voice samples only, with no further information about the situation or about the patient. The addition of the "new features" improved the classification results.

Our results comply with previous studies which found it possible to distinguish "no pain" from "pain" using various physiological indications, but had difficulty in differentiating between the levels of pain [12]. A possible explanation for this finding might be the subjectivity of the experience of pain, and the significant dissimilarity in the ways different people assess their pain [1]. This dissimilarity is very dominant in fine classification, but may be resolved in two-category classification.

Our results were achieved with voice samples that were recorded in the patient's natural environment. This fact had some disadvantages, but indicates that future practical use of the findings might be done in medical and therapeutic areas, and not confined to "sterile" surroundings only.

Further research can consider information regarding medical treatment. Another direction that can be examined is weather the connection between pain and voice is direct or indirect. For example, the reported pain might have been caused by physiological or mental effects, which in turn might have caused the vocal change which we found. For example, stress, anxiety or fear might be the real cause of the effects on the voice, rather than the pain itself. In this paper, we consider the final result of the pain, neglecting the direct source or the causes.

This is a preliminary proof of concept research. We hope that the capability to distinguish significant pain from no significant pain might be useful in practical uses, and especially as an aide to clinicians in more objectively assessing the pain which their patients experience.

## 7. References


[1] J. J. Bonica, "The need of a taxonomy," *Pain*, no. 6(3), pp. 247-248, 1979.

[2] M. Loggia, M. Juneau and M. Bushnell, "Autonomic responses to heat pain: heart rate, skin conductance, and their relation to verbal ratings and stimulus intensity," *Pain*, no. 152, pp. 592-8, 2011.

[3] A. Moeltner, R. Hoelzl and F. Strian, "Heart rate changes as an autonomic component of the pain response," *Pain*, vol. 43, no. 1, pp. 81-89, Oct 1990.

[4] B. Robinson, "Relation of Heart Rate and Systolic Blood Pressure to the Onset of Pain in Angina Pectoris," *Circulation Research*, no. 35, pp. 1073-1083, 1967.

[5] R. Orlikoff and R. Baker, "The effect of the heartbeat on vocal fundamental frequency perturbation," *Journal of Speech and Hearing Research,* no. 32, pp. 576-582, 1989.

[6] R. Orlikoff, "Vowel amplitude variation associated with the heart cycle," *Journal of Acoustical Society of America,* no. 88, pp. 2091-2098, 1990.

[7] C. Giddens, K. Barron, J. Byrd-Craven, K. Clark and A. Winter, "Vocal Indices of Stress: A Review," *Journal of Voice,* no. 27, pp. 390-398, 2013.

[8] B. Schuller, S. Steidl, A. Batliner, F. Burkhardt, L. Devillers, C. Müller and S. Narayanan, "Paralinguistics in speech and language - State-of-the-art and the challenge," *Computer Speech and Language,* no. 27, pp. 4-39, 2013.

[9] S. Alghowinem, R. Goecke, M. Wagner, J. Epps, M. Breakspear and G. Parker, "From Joyous to Clinically Depressed: Mood Detection Using Spontaneous Speech," *In FLAIRS Conference,* May 2012.

[10] S. Alghowinem, R. Goecke, M. Wagner, J. Epps, M. Breakspear and G. Parker, "Detecting depression: A comparison between spontaneous and read speech," *In Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on,* pp. 7547-7551, May 2013.

[11] M. ElAyadi, S. Kamel and F. Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases," *Pattern Recognition,* no. 44, pp. 572-587, 2011.

[12] R. Treister, M. Kliger, G. Zuckerman, I. Goor Aryeh and E. Eisenberg, "Differentiating between heat pain intensities: the combined effect of multiple autonomic parameters.," *pain,* 2012.

[13] F. Eyben, M. Woellmer and B. Schuller, "Opensmile: the munich versatile and fast open-source audio feature extractor," *In Proc. ACM Multimedia (MM),* pp. 1459-1462, Oct. 2010.

[14] F. Eyben, M. Woellmer and B. Schuller, "The openSMILE book - openSMILE: The Munich Versatile and Fast Open-Source Audio Feature Extractor," 23 may 2010. [Online]. Available: http://opensmile.sourceforge.net/. [Accessed 23 Apr 2014].

[15] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and I. H. Witten, "The WEKA Data Mining Software: An Update," *SIGKDD Explorations,* vol. 11, no. 1, 2009.

[16] M. A. Hall, *Correlation-based Feature Selection for Machine Learning,* Hamilton: The University of Waikato, 1999.

[17] S. Taimela, K. Osterman, H. Alaranta, A. Soukka and U. Kujala, "Long psychomotor reaction time in patients with chronic low-back pain: preliminary report," *Archives of Physical Medicine and Rehabilitation,* vol. 74, no. 11, pp. 1161-1164, 1993.

[18] A. Reinersmanna, G. S. Haarmeyer, M. Blankenburg, J. Frettlöh, E. K. Krumova, S. Ocklenburg and C. Maier, "Left is where the L is right. Significantly delayed reaction time in limb laterality recognition in both CRPS and phantom limb pain patients," *Neuroscience Letters,* vol. 486, no. 3, pp. 240-245, 2010.

[19] G. Crombez, C. Eccleston, F. Baeyens and P. Eelen, "The disruptive nature of pain: An experimental investigation," *Behaviour Research and Therapy,* vol. 34, no. 11-12, pp. 911-918, 1996.

[20] H. Hermansky, "Perceptual linear predictive (PLP) analysis of speech," *the Journal of the Acoustical Society of America,* vol. 87, no. 4, pp. 1738-1752, 1990.

[21] D. O'Shaughnessy, "Linear predictive coding," *IEEE Potentials,* vol. 7, no. 1, pp. 29-32, 1988.

[22] H. Hermansky, N. Morgan, A. Bayya and P. Kohn, "RASTA-PLP Speech Analysis Technique," *Acoustics, Speech, and Signal Processing, IEEE International Conference on,* vol. 1, pp. 121-124, 1992.

[23] B. Gerazov and Z. Ivanovski, "Overview of Feature Selection for Automatic Speech Recognition," *Audio Engineering Society Convention 132,* Apr 2012.

[24] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and Psychological Measurement,* vol. 20, no. 1, pp. 37-46, 1960.

[25] J. Carletta, "Assessing agreement on classification tasks: The kappa statistic," *Coputational Linguistics,* vol. 22, no. 2, pp. 249-254, 1996.

[26] I. H. Witten, E. Frank and M. A. Hall, Data mining, Practical Machine Learning Tools and Techniques, 3rd ed., Burlington MA: Morgam Kaufmann, 2011.

[27] B. Schuller, A. Batliner, S. Steidl and D. Seppi, "Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge," *Speech Communication,* vol. 53, no. 9-10, p. 1062–1087, 2011.

[28] J. C. Platt, "Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines," 1998.