Psychiatria Danubina, 2013; Vol. 25, No. 3, pp 340-346
© Medicinska naklada - Zagreb, Croatia

Conference paper

COMPUTER-AIDED PSYCHOTHERAPY BASED ON MULTIMODAL ELICITATION, ESTIMATION AND REGULATION OF EMOTION

Krešimir Ćosić[1], Siniša Popović[1], Marko Horvat[1], Davor Kukolja[1], Branimir Dropuljić[1], Bernard Kovač[1] & Miro Jakovljević[2]

[1]University of Zagreb Faculty of Electrical Engineering and Computing, Croatia
[2]University Hospital Centre Zagreb, Department of Psychiatry, Croatia

SUMMARY

Contemporary psychiatry is looking at affective sciences to understand human behavior, cognition and the mind in health and disease. Since it has been recognized that emotions have a pivotal role for the human mind, an ever increasing number of laboratories and research centers are interested in affective sciences, affective neuroscience, affective psychology and affective psychopathology. Therefore, this paper presents multidisciplinary research results of Laboratory for Interactive Simulation System at Faculty of Electrical Engineering and Computing, University of Zagreb in the stress resilience. Patient's distortion in emotional processing of multimodal input stimuli is predominantly consequence of his/her cognitive deficit which is result of their individual mental health disorders. These emotional distortions in patient's multimodal physiological, facial, acoustic, and linguistic features related to presented stimulation can be used as indicator of patient's mental illness. Real-time processing and analysis of patient's multimodal response related to annotated input stimuli is based on appropriate machine learning methods from computer science. Comprehensive longitudinal multimodal analysis of patient's emotion, mood, feelings, attention, motivation, decision-making, and working memory in synchronization with multimodal stimuli provides extremely valuable big database for data mining, machine learning and machine reasoning. Presented multimedia stimuli sequence includes personalized images, movies and sounds, as well as semantically congruent narratives. Simultaneously, with stimuli presentation patient provides subjective emotional ratings of presented stimuli in terms of subjective units of discomfort/distress, discrete emotions, or valence and arousal. These subjective emotional ratings of input stimuli and corresponding physiological, speech, and facial output features provides enough information for evaluation of patient's cognitive appraisal deficit. Aggregated real-time visualization of this information provides valuable assistance in patient mental state diagnostics enabling therapist deeper and broader insights into dynamics and progress of the psychotherapy.

Key words: *emotion elicitation - multimodal stimulation - cognitive appraisal - multimodal features - emotion estimation – physiological – acoustic – linguistic – facial features*

* * * * *

INTRODUCTION

Multidisciplinary research in affective neuroscience and affective computing supported by technological innovations are changing contemporary psychotherapy (Sander 2013). Concept proposed in this paper attempts to delegate routine tasks to computer-aided tools and means. Furthermore, available computer-aided tools and means can automatically monitor and track patient's therapeutic progress on the basis of comprehensive multimodal signal analysis. Proposed concept is based on multimodal emotion elicitation and estimation of patients' emotion and mood variability within session, as well as across all sessions. Patient's awareness of better emotion regulation during psychotherapy time course should strengthen his/her cognitive appraisal and cognitive regulatory mechanisms. Increasing awareness of uncontrollable emotional behavior and potential of cognitive control of internal thoughts, sensations, and emotions is important for cognitively based stress management techniques. The training and learning of appropriate coping strategies that improve and maintain patient's emotional stability in stressful situations are critically important for resilience building. Better regulation of induced negative emotions may reduce mood fluctuations and variability due to patients' faster and better control of their thoughts what can make them less anxious and healthier (Cloitre et al. 2002, Rodebaugh & Heimberg 2008). Such real-time regulatory cognitive process based on patient's enhanced ability to control their thoughts, and their more rational interpretation of trauma-related stimuli has enormous positive impact on patient's mental health. This cognitive restructuring may help patients to become more aware of interdependency between their thoughts and their autonomic (ANS) and central nervous system (CNS) response, and intends to modify their distorted thoughts whenever they arise. Such technologically assisted patient's cognitive restructuring based on higher awareness and better understanding of interdependency between traumatic stimuli semantics and context and corresponding multimodal features will be extremely important in modern psychotherapy. Modification of patient's cognitive appraisal distortions can be supported by more effective real-time closed-loop stimulation strategies led by therapist and proposed computer-aided tools (Ćosić et al. 2010). Modifying maladaptive cognitions by "brain-train" exercise that reduces highly aroused brain regions using proposed tools and means is of considerable importance in trea0ting many mental disorders.

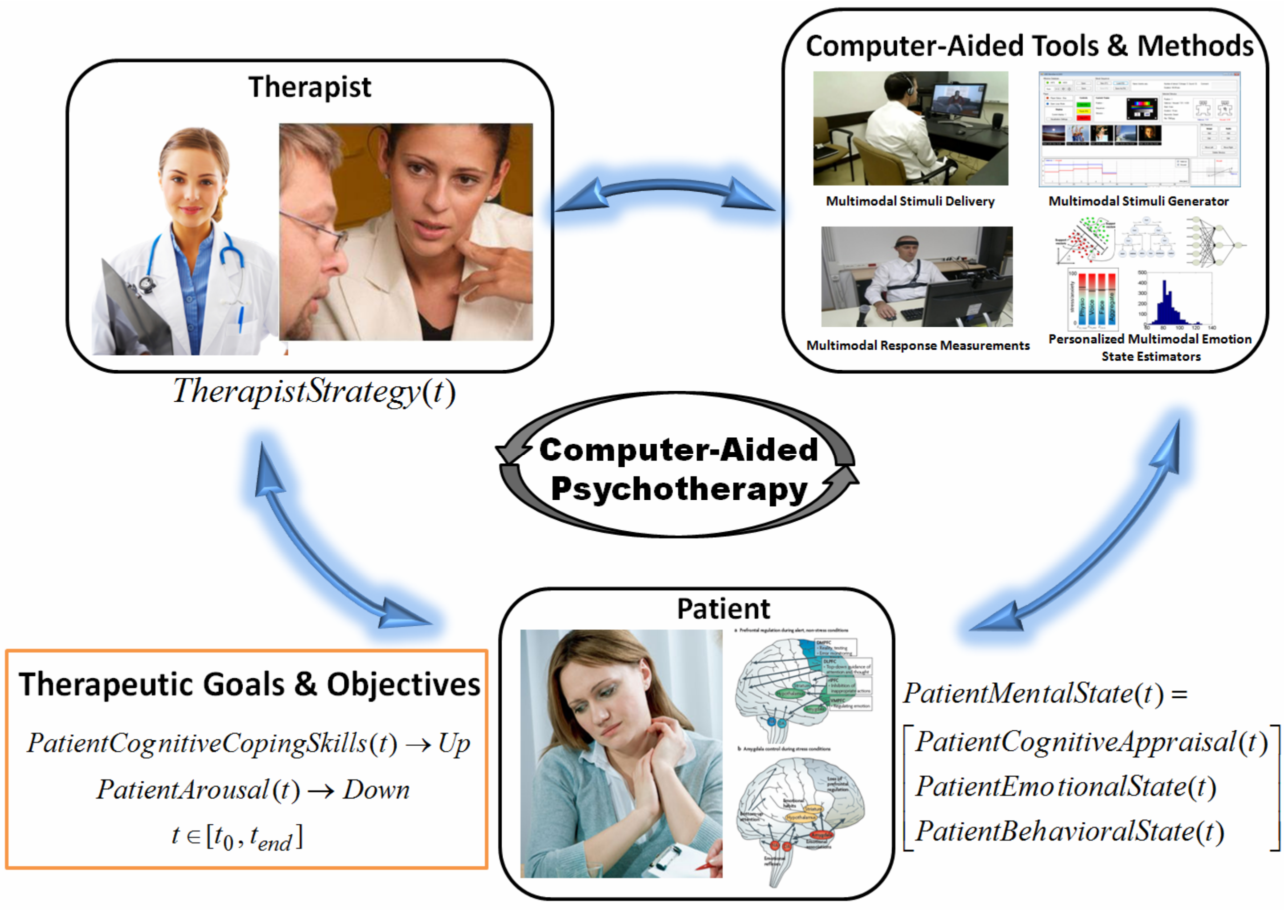

Therapist
Computer-Aided Tools & Methods
Multimodal Stimuli Delivery
Multimodal Stimuli Generator
Multimodal Response Measurements
Personalized Multimodal Emotion State Estimators
$TherapistStrategy(t)$
Computer-Aided Psychotherapy
Patient
Therapeutic Goals & Objectives
$PatientCognitiveCopingSkills(t) \rightarrow Up$
$PatientArousal(t) \rightarrow Down$
$t \in [t_0, t_{end}]$
$PatientMentalState(t) = \begin{bmatrix} PatientCognitiveAppraisal(t) \\ PatientEmotionalState(t) \\ PatientBehavioralState(t) \end{bmatrix}$

Figure 1. Multimodal dynamic interactions between therapist and patients in computer-aided psychotherapy

The concept proposed in this paper and shown in Figure 1 is based on a dynamic interplay between the therapist, patient and computer-aided tools and methods for multimodal emotion elicitation and estimation. These computer-aided tools and methods are related to personalized multimodal stimuli generation, multimodal response measurement and emotion estimation.

One typical session of proposed protocol is illustrated by Figure 2. The therapist delivers a variety of audio-visual stimuli while interacting with the patient via appropriate therapeutic instructions, questions, clarifications, narratives etc., using tools for multimodal elicitation of emotion. More details regarding elicitation and estimation of the patient's emotions are provided in subsequent sections.

The brain area that plays a key role in proposed psychotherapeutic approach is mainly focused on prefrontal cortex, i.e. a more effective "top-down" regulation of hyperexcitable limbic structures by prefrontal control systems, i.e. on more prefrontal cortex activation and amygdala-hippocampal deactivation (Gross 2007). It is also well known from neurobiological standpoint that stress impairs prefrontal cortex connections that are responsible for complex cognitive functions. It means that stress ruins complex cognitive abilities (Arnsten 2009, Seung 2012) and therefore, learning and strengthening of cognitive self-regulation of emotion is extremely important in the context of stress resistance and stress resilience. Effectiveness of emotion self-regulation skills depends on prefrontal cortex and amygdala interactions, i.e. interactions between cognitive and emotional brain (Salzman & Fusi 2010), as well as on other brain regions involved in bottom-up and top-down emotion regulation (Ochsner et al. 2009, McRae et al. 2012). Rumination of stressful and traumatic events, objects and situations during emotion elicitation requires well-synchronized and timely therapist intervention, focused on strengthening the activation of patient's orbitofrontal cortex and ventromedial prefrontal cortex (OFC/vmPFC), i.e. on the cognitive inhibition of the patient's amygdala (Quirk & Gehlert 2003). Such psychotherapeutic sessions mainly focused on fear extinction may produce specific synaptic reinforcement which strengthens the inhibitory connections from the OFC/vmPFC to the amygdala (Sotres-Bayon et al. 2006, Akirav & Maroun 2007, Davis 2011). Therefore, more neurobiological research efforts should be focused on better understanding of positive change induced by such technologically supported emotion regulation techniques.

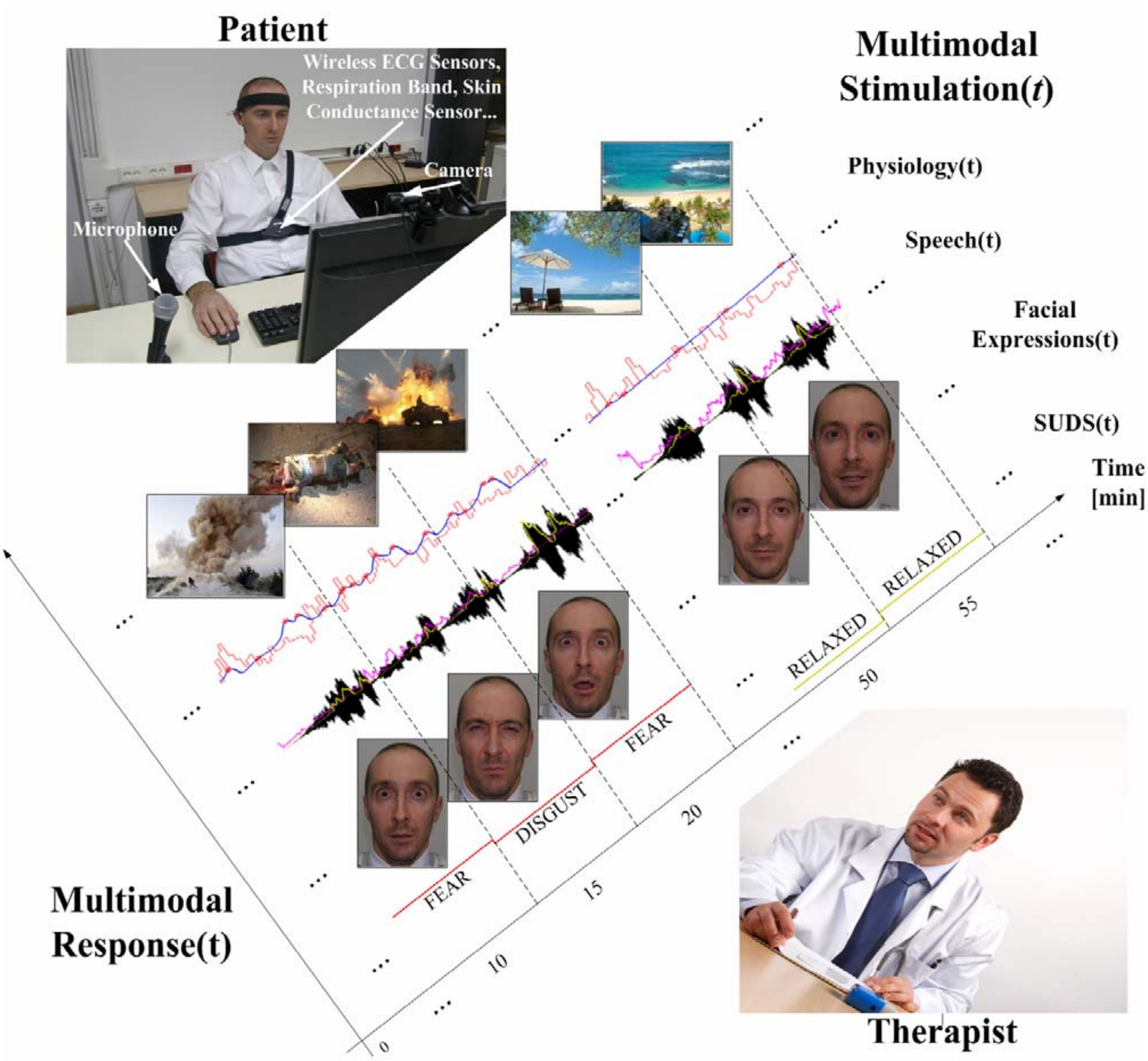


Figure 2. Illustration of time-synchronized stimuli delivery and patient's response measurement during typical computer aided psychotherapy session

MULTIMODAL PERSONALIZED EMOTION ELICITATION

Multimodal elicitation of emotion combines computer multimedia e.g. pictures, natural and manmade sounds, recorded speech, written text messages, video-clips and films or virtual reality synthetic environments, with spoken narratives, as well as face-to-face conversation between the therapist and the patient. The goal of multimodal elicitation is to provide optimal personalized stimulations that support cognitive restructuring along series of psychotherapeutic sessions and enhance effectiveness of therapy. In order to achieve the necessary richness of input stimuli (Mareels 1984) during the course of psychotherapeutic sessions, the therapist engages the patient into a structured stimuli-related conversation using questions like "please describe what you are seeing", "does this remind you on some particular situation in your life", etc. as well as any other specific questions that help to provoke the patient's emotions and emotion-related states. Initial psychotherapeutic session is typically related to the patient's exposure to a variety of stimuli intended to identify the context of traumatic experience based on his/her physiological, facial, acoustic, and linguistic reactions.

The stimuli personalization is based on appropriate questionnaires and interviews related to the patient's stressful experiences and stressful situations encountered in the past. Such questionnaires may include Life Events Checklist (Gray et al. 2004), Life Experience Survey (Sarason et al. 1978), Psychiatric Epidemiological Research Inventory (Dohrenwend et al. 1978), or any other standard or custom-made questionnaires regarding patient's stressful events encountered in life. Such approach facilitates selection of personalized multimodal stimuli, like specific images, video clips, narrative, as well as questions raised by therapist that correspond to the patient's emotional traumatic experience, etc. It may also include augmented virtual reality which integrates multimedia with state-of-the-art real-time computer graphics to achieve higher degree of immersion. Virtual environments may simulate visual and audio types of multimedia formats, and can be supplemented with real-life records to create new more powerful stimuli with broader and higher elicitation of associative cortex. But personalized real-life multimedia are the most realistic stimuli with the highest appraisal potential (Coan & Allen 2007).

Comprehensive emotion elicitation tools illustrated by Figure 3 enable psychiatrists and psychologists to perform intuitive retrieval of a variety of audio-visual stimuli from semantically and emotionally annotated stimuli databases.

MULTIMODAL PERSONALIZED EMOTION ESTIMATION BASED ON PHYSIOLOGICAL, ACOUSTIC AND FACIAL SIGNALS

New low cost micro-sensor technologies enable measurements and monitoring of the patient's multimodal physiological, acoustic, linguistic, and facial reactions (Ćosić et al. 2012), which can detect a variety of invisible nonverbal and verbal cues of patients during the therapy, what is almost impossible even for the most experienced therapists. Such invisible patient's traits and cues along therapeutic sessions might be extremely valuable in psychotherapy, avoiding potential misinterpretation and even misunderstandings of patient's reactions during a series of therapeutic sessions. Computer can track each session, its timing, duration, content, total number and mean duration of all sessions, the patient's self-ratings along session, as well as his/her multimodal emotional response in real time, and stores all these data into personalized patient's database. Real-time comparative analysis of multimodal stimuli and multimodal response based on physiological, speech, as well as facial modalities enable therapist to have better and deeper real-time comprehensive insight into variation and fluctuation of patient's mental state during such session protocols.

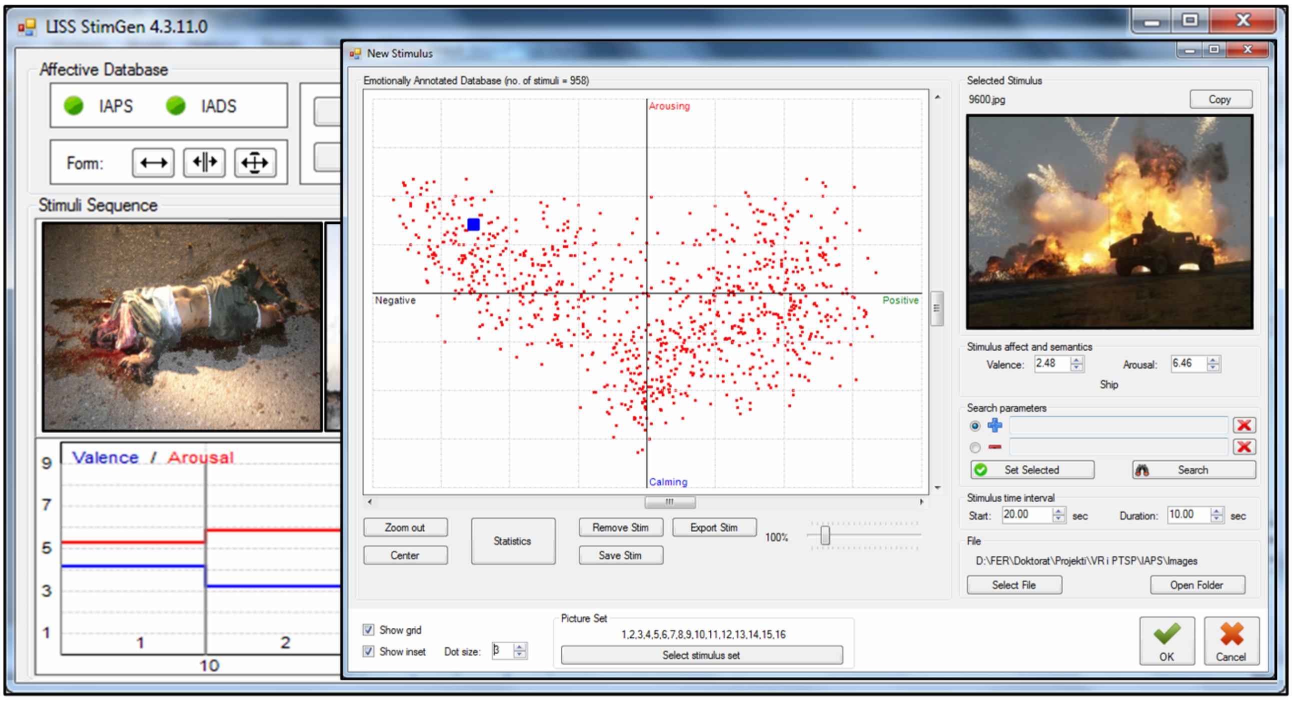
Proposed multimodal personalized emotion estimation takes into account individual differences among patients, differences in their mental disorders signatures and patterns, differences in their neurobiological and anatomical networks, as well as differences due to their multimodal longitudinal individual variation, i.e. fluctuations within session and from session to session. Variety of stimuli-induced features could highly correlate with patient's specific mental disorders. Such features, like mean value, standard deviation, latency, rise time, spectral bandwidth, increased volume of speech, changes in facial expression, other forms of motor expression, could provide huge amount of information about patient clinical progress, more information about critical turning point during psychotherapeutic sessions etc. what might be helpful in identifying patient's specific mental problems.

Physiological signals

There is good evidence in psychophysiology that specific physiological activities are associated with related affective states (Bradley 2000). Indicators of peripheral physiological activity, such as cardiovascular activity, electrodermal activity, and electromyographic (EMG) activity, have long been used even as primary indicators of emotion (Lang 1979, Lazarus 1968).

Physiological signals, like skin conductance, electrocardiogram (ECG), heart rate, respiration rate etc., are dominantly related to the ANS activity and conscious manipulation of these signals is much harder than speech or facial emotional expressions (Kim & André 2008). From these signals a wide range of physiological features can be computed in time/frequency, entropy, geometric analysis, sub-band spectra, multi scale entropy domains etc. Typical physiological features are statistical measures such as mean, standard deviation, minimum, maximum or mean of first difference of each signal. Some additional features can be based on heart rate variability measures (Camm et al. 1996), skin conductance response measures (Boucsein 2011) etc.

Finding the dominant set of physiological features, that are the most relevant for differentiating various mental disorders is one of the main objectives of emotional state estimation based on physiology. Statistical results demonstrate that most dominant physiological features for example for emotional state of fear are related with skin conductance response (SCR): peak-to-peak amplitude and standard deviation of skin conductance signal (Lang 1995, Healey 2000). SCR varies linearly with arousal ratings (Lang 1995) and has been used as a measure of stress in anticipatory anxiety studies including studies of public speaking (Healey 2000). Computed physiological features are inputs to the emotional state estimator that transforms these inputs into the emotional state of the patient. State estimator can be obtained by using, for example, an artificial neural network (Haykin 1999), support vector machine (Cortes & Vapnik 1995), decision tree (Witten et al. 2011) etc.



Emotionally Annotated Database (no. of stimuli = 958)

Selected Stimulus: 9600.jpg

Stimulus affect and semantics: Valence: 2.48 Arousal: 6.46 Ship

Search parameters: ● ○

Stimulus time interval: Start: 20.00 sec Duration: 10.00 sec

File: D:\FER\Doktorat\Projekti\VR i PTSP\IAPS\Images

☑ Show grid

☑ Show inset Dot size: 3

Picture Set: 1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16

Figure 3. New tools for multimodal emotion elicitation

Acoustic signals

Real-time analysis of patient speech features provides extremely valuable information about dynamics of patient's emotional change during the psychotherapeutic sessions. Based on real-time analysis of patient's speech i.e. acoustic and linguistic features, patient's emotional variations and fluctuations can be accurately estimated, what is in some cases almost impossible to be observed by an average listener. Parameters of speech-based emotion estimation algorithms are iteratively adapted within and after each session based on individual patient's input-output transfer function related to annotated input stimulus and corresponding speech features.

Acoustic features that contain relevant emotional information can be calculated from non-verbal cues in a patient's utterance, including prosodic parameters like the vocal cords oscillation frequency (pitch), short-term energy contour of the speech, zero cross rate and the speech spectrum distribution (Banse & Scherer 1996, Schuller et al. 2004, Rong et al. 2009, Yun & Yoo 2012). Typical acoustic features are related to some statistical measures across the utterance like mean value, standard deviation, minimum, maximum, etc. Some additional features can be based on time duration of voiced/unvoiced/silence segments in an utterance (Schuller et al. 2004), perturbation or irregularity of pitch contour (jitter) (Banse & Scherer 1996) and speech amplitude modulation (shimmer) (Li et al. 2007) etc. Features calculated from jitter and shimmer can indicate certain patient's mental disorders like anxiety (Fuller et al. 1992), as well as patient's current emotional state e.g. fear (Scherer 1986). Therefore, these jitter and/or shimmer related features may be used as dominant features in phobia, panic disorder or social anxiety disorder treatments.

Facial expression

Many studies illustrate the relationship between the facial expressions and various psychiatric disorders, like recognition of schizophrenia and depression which are characterized by reduced facial activity especially in the upper face (Gaebel & Wolwer 2004). Facial expressions are shaped through muscular activity which is driven by complex neural control networks that includes both autonomous and voluntary components and different brain structures mainly in occipital and temporal lobes (Adolphs 2002).

The correspondence between the subjective emotional ratings and the presence, or absence, of specific Action Units (Ekman & Friesen 1978) which characterize reported emotional state, can indicate the existence of certain patient's mental disorder. Using different computer vision methods, it is possible to locate and extrapolate the face with its spatial characteristics of main anatomical regions from the given video or image data, e.g. by using Active Shape Models (Cootes et al. 1995). Automated emotion recognition from facial expressions can be conducted by several algorithms for pattern recognition, like support vector machines (SVMs), artificial neural networks etc. Given that the reduction or absence of facial activity can be an indicator of certain mental disorder (Gaebel & Wolwer 2004), it is very important to objectively measure and evaluate the overall patient's facial activity during the emotion elicitation process throughout the therapy sessions. There are a variety of options for evaluating this mental disorder-related facial feature, like the sum of all relative spatial offsets of facial characteristic points, e.g. ASM characteristic points, from the baseline neutral face point set. The differences between the upper and lower face activity can be evaluated using the same approach limited to particular subset of facial characteristic points that represent certain morphological structure belonging to the upper or lower part of the face. Information regarding differences between the estimation result of patient's personalized estimator and the expected emotion estimation value, e.g. by using referent estimator learned over the general healthy population, can also be integrated in diagnostic process of patient's mental disorders.

CONCLUSION

This paper presents computer-aided psychotherapy based on real time interactive multimodal elicitation and estimation of emotion, which may significantly improve effectiveness of traditional face-to-face psychotherapy. Emotion elicitation is based on generation of personalized therapy-relevant audio-visual stimuli, while real-time emotion estimation is related to patient's physiology, speech, and facial signals, as well as corresponding features. Such concept of technologically assisted psychotherapeutic cognitive restructuring may support resilience of patients suffering from major depressive disorder, obsessive-compulsive disorder, panic disorder, social anxiety disorder, specific phobias, and even posttraumatic stress disorder. Comparative analysis of stimuli versus multimodal response features, which is facilitated by the proposed computer-aided tools and means, could also reveal specific correlations between elicited stimuli and specific dominant features that discriminate patients from healthy people. Furthermore, future research of multimodal features and their correlations with a variety of psychopathologies will be extremely important for further development of proposed computer-aided psychotherapeutic programs and their evaluation metrics. In summary, more future interdisciplinary efforts among psychiatrists, psychologists, neuroscientists, engineers and computer scientists should enhance proposed state-of-the-art analytical tools and means which can open new niche in personalized mental health medicine.

Acknowledgements

This research has been partially supported by the Ministry of Science, Education and Sports of the Republic of Croatia.

Conflict of interest: None to declare.

References

1. *Adolphs R: Recognizing emotion from facial expressions: psychological and neurological mechanisms. Behav Cogn Neurosci Rev 2002; 1:21-61.*
2. *Akirav I & Maroun M: The role of the medial prefrontal cortex-amygdala circuit in stress effects on the extinction of fear. Neural Plast 2007; 2007:30873. doi:10.1155/2007/30873.*
3. *Arnsten AFT: Stress signalling pathways that impair prefrontal cortex structure and function. Nature Reviews Neuroscience 2009; 10:410-22.*
4. *Banse R & Scherer K: Acoustic profiles in vocal emotion expression. J Pers Soc Psychol 1996; 70:614-36.*
5. *Boucsein W: Electrodermal Activity. Springer Verlag, 2011.*
6. *Bradley MM: Emotion and motivation. In Cacioppo JT, Tassinary LG & Berntson G (eds): Handbook of Psychophysiology, 602-42. Cambridge University Press, New York, 2000.*
7. *Camm AJ, Malik M, Bigger JT, Breithardt G, Cerutti S, Cohen RJ et al: Heart rate variability: standards of measurement, physiological interpretation, and clinical use. Circulation 1996; 93:1043-65.*
8. *Cloitre M, Koenen KC, Cohen LR & Han H: Skills training in affective and interpersonal regulation followed by exposure: a phase-based treatment for PTSD related to childhood abuse. J Consult Clin Psychol 2002; 70:1067-74.*
9. *Coan JA & Allen JJB: Handbook of Emotion Elicitation and Assessment. Oxford University Press, New York, 2007.*
10. *Cootes TF, Cooper D, Taylor CJ & Graham J: Active Shape Models – their training and application. Computer Vision and Image Understanding 1995; 61:38-59.*
11. *Cortes C & Vapnik V: Support-vector networks. Machine Learning 1995; 20:273-97.*
12. *Ćosić K, Popović S, Kukolja D, Horvat M & Dropuljić B: Physiology-driven adaptive virtual reality stimulation for prevention and treatment of stress related disorders. Cyberpsychology, Behavior, and Social Networking 2010; 13:73-8.*
13. *Ćosić K, Popović S, Horvat M, Kukolja D, Dropuljić B, Kovač B et al: Multimodal paradigm for mental readiness training and PTSD prevention. Paper presented at NATO Advanced Study Institute on Invisible Wounds – New Tools to Enhance PTSD Diagnosis and Treatment, 2012 Jun 18-28, Ankara, Turkey.*
14. *Davis M: NMDA receptors and fear extinction: implications for cognitive behavioral therapy. Dialogues in Clinical Neuroscience 2011; 13:463-74.*
15. *Dohrenwend BS, Askenasy AR, Krasnoff L & Dohrenwend BP: Exemplification of a Method for Scaling Life Events: The PERI Life Events Scale. J Health Soc Behav 1978; 19:205-29.*
16. *Ekman P & Friesen W: Facial Action Coding System: a technique for the measurement of facial movement. Consulting Psychologists Press, Palo Alto, CA, 1978.*
17. *Fuller BF, Horii Y & Conner DA: Validity and reliability of nonverbal voice measures as indicators of stressor-provoked anxiety. Research in Nurse & Health 1992; 15:379-89.*
18. *Gaebel W & Wolwer W: Facial expressivity in the course of schizophrenia and depression. Eur Arch Psychiatry Clin Neurosci 2004; 254:335-42.*

19. Gray MJ, Litz BT, Hsu JL & Lombardo TW: The psychometric properties of the Life Events Checklist. Assessment 2004; 11:330-41.
20. Gross JJ: Handbook of Emotion Regulation. Guilford Press, New York, NY, 2007.
21. Haykin S: Neural Networks: A Comprehensive Foundation. Prentice Hall, 1999.
22. Healey JA: Wearable and automotive systems for affect recognition from physiology. PhD thesis. Massachusetts Institute of Technology, 2000.
23. Kim J & André E: Emotion recognition based on physiological changes in music listening. IEEE Transaction on Pattern Analysis and Machine Intelligence 2008; 30:2067-83.
24. Lang PJ: A bio-informational theory of emotional imagery. Psychophysiology 1979; 16:495-512.
25. Lang PJ: The emotion probe: studies of motivation and attention. Am Psychol 1995; 50:372-85.
26. Lazarus RS: Emotions and adaptation: conceptual and empirical relations. In Arnold WJ (ed): Proceedings of the Nebraska Symposium on Motivation, 175-266. University of Nebraska Press, Lincoln, 1968.
27. Li X, Tao J, Johnson MT, Soltis J, Savage A, Leong KM et al.: Stress and emotion classification using jitter and shimmer features. Proceedings ICASSP ’07, 1081-4. 2007.
28. Mareels I: Sufficiency of excitation. Systems & Control Letters 1984; 5:159-63.
29. McRae K, Misra S, Prasad AK, Pereira SC & Gross JJ: Bottom-up and top-down emotion generation: implications for emotion regulation. Social, Cognitive & Affective Neuroscience 2012; 7:253-62.
30. Ochsner KN, Ray RR, Hughes B, McRae K, Cooper JC, Weber J et al.: Bottom-up and top-down processes in emotion generation: common and distinct neural mechanisms. Psychological Science 2009; 20:1322-31.
31. Quirk GJ & Gehlert DR: Inhibition of the amygdala: key to pathological states? Ann N Y Acad Sci 2003; 985:263-72.
32. Rodebaugh TL & Heimberg RG: Emotion regulation and the anxiety disorders: adopting a self-regulation perspective. In Vingerhoets A, Nyklicek I & Denollet J (eds): Emotion Regulation: Conceptual and Clinical Issues, 140-49. Springer, 2008.
33. Rong J, Li G & Chen YP: Acoustic feature selection for automatic emotion recognition from speech. Information Processing and Management 2009; 45:315-28.
34. Salzman CD & Fusi S: Emotion, cognition, and mental state representation in amygdala and prefrontal cortex. Annu Rev Neurosci 2010; 33:173-202.
35. Sander D: Preface: The power of emotions. In Sander D (ed): The Power of Emotions, 7-12. Belin: Editeur Independent, Depuis 1777, Paris, 2013.
36. Sarason IG, Johnson JH & Siegel JM: Assessing the impact of life changes: development of the Life Experiences Survey. J Consult Clin Psychol 1978; 46:932-46.
37. Scherer K: Vocal affect expression: a review and a model for future research. Psychol Bull 1986; 99:143-65.
38. Schuller B, Rigoll G & Lang M: Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine-belief network architecture. Proceedings of ICASSP ’04, 577-80. 2004.
39. Seung S: Connectome: How the Brain's Wiring Makes Us Who We Are. Houghton Mifflin Harcourt Trade, 2012.
40. Sotres-Bayon F, Cain CK & LeDoux JE: Brain mechanisms of fear extinction: historical perspectives on the contribution of prefrontal cortex. Biol Psychiatry 2006; 60:329-36.
41. Witten IH, Frank E & Hall MA: Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, 2011.
42. Yun S & Yoo CD: Loss-scaled large-margin gaussian mixture models for speech emotion classification. IEEE Transactions on Audio, Speech and Language Processing 2012; 20:585-98.

Correspondence:
Siniša Popović
University of Zagreb Faculty of Electrical Engineering and Computing
Unska 3, 10000 Zagreb, Croatia
E-mail: sinisa.popovic@fer.hr