

World Journal of *Psychiatry*

World J Psychiatry 2023 January 19; 13(1): 1-35



REVIEW

- 1 Emotion recognition support system: Where physicians and psychiatrists meet linguists and data engineers

Adibi P, Kalani S, Zahabi SJ, Asadi H, Bakhtiar M, Heidarpour MR, Roohafza H, Shahooh H, Amouzadeh M

ORIGINAL ARTICLE**Clinical and Translational Research**

- 15 Evaluation of a culturally adapted cognitive behavior therapy-based, third-wave therapy manual

Phiri P, Clarke I, Baxter L, Zeng YT, Shi JQ, Tang XY, Rathod S, Soomro MG, Delanerolle G, Naeem F

ABOUT COVER

Editorial Board Member of *World Journal of Psychiatry*, Sujita Kumar Kar, MD (Psychiatry), Additional Professor, Department of Psychiatry, King George's Medical University, Lucknow-226003, UP, India. drsujita@gmail.com

AIMS AND SCOPE

The primary aim of *World Journal of Psychiatry* (WJP, *World J Psychiatry*) is to provide scholars and readers from various fields of psychiatry with a platform to publish high-quality basic and clinical research articles and communicate their research findings online.

WJP mainly publishes articles reporting research results and findings obtained in the field of psychiatry and covering a wide range of topics including adolescent psychiatry, biological psychiatry, child psychiatry, community psychiatry, ethnopsychology, psychoanalysis, psychosomatic medicine, etc.

INDEXING/ABSTRACTING

The WJP is now abstracted and indexed in Science Citation Index Expanded (SCIE, also known as SciSearch®), Current Contents/Clinical Medicine, Journal Citation Reports/Science Edition, PubMed, PubMed Central, Reference Citation Analysis, China National Knowledge Infrastructure, China Science and Technology Journal Database, and Superstar Journals Database. The 2022 Edition of Journal Citation Reports® cites the 2021 impact factor (IF) for WJP as 3.500; IF without journal self cites: 3.313; 5-year IF: 7.380; Journal Citation Indicator: 0.62; Ranking: 89 among 155 journals in psychiatry; and Quartile category: Q3.

RESPONSIBLE EDITORS FOR THIS ISSUE

Production Editor: *Yu-Xi Chen*; Production Department Director: *Xu Guo*; Editorial Office Director: *Yun-Xiao Jiao Wu*.

NAME OF JOURNAL

World Journal of Psychiatry

ISSN

ISSN 2220-3206 (online)

LAUNCH DATE

December 31, 2011

FREQUENCY

Monthly

EDITORS-IN-CHIEF

Rajesh R Tampi, Ting-Shao Zhu, Panteleimon Giannakopoulos

EDITORIAL BOARD MEMBERS

<https://www.wjgnet.com/2220-3206/editorialboard.htm>

PUBLICATION DATE

January 19, 2023

COPYRIGHT

© 2023 Baishideng Publishing Group Inc

INSTRUCTIONS TO AUTHORS

<https://www.wjgnet.com/bpg/gerinfo/204>

GUIDELINES FOR ETHICS DOCUMENTS

<https://www.wjgnet.com/bpg/GerInfo/287>

GUIDELINES FOR NON-NATIVE SPEAKERS OF ENGLISH

<https://www.wjgnet.com/bpg/gerinfo/240>

PUBLICATION ETHICS

<https://www.wjgnet.com/bpg/GerInfo/288>

PUBLICATION MISCONDUCT

<https://www.wjgnet.com/bpg/gerinfo/208>

ARTICLE PROCESSING CHARGE

<https://www.wjgnet.com/bpg/gerinfo/242>

STEPS FOR SUBMITTING MANUSCRIPTS

<https://www.wjgnet.com/bpg/GerInfo/239>

ONLINE SUBMISSION

<https://www.f6publishing.com>



Emotion recognition support system: Where physicians and psychiatrists meet linguists and data engineers

Peyman Adibi, Simindokht Kalani, Sayed Jalal Zahabi, Homa Asadi, Mohsen Bakhtiar, Mohammad Reza Heidarpour, Hamidreza Roohafza, Hassan Shahoon, Mohammad Amouzadeh

Specialty type: Psychiatry

Provenance and peer review:

Invited article; Externally peer reviewed.

Peer-review model: Single blind

Peer-review report's scientific quality classification

Grade A (Excellent): A

Grade B (Very good): B

Grade C (Good): 0

Grade D (Fair): 0

Grade E (Poor): 0

P-Reviewer: Panduro A, Mexico;
Stoyanov D, Bulgaria

Received: June 19, 2022

Peer-review started: June 19, 2022

First decision: September 4, 2022

Revised: September 18, 2022

Accepted: December 21, 2022

Article in press: December 21, 2022

Published online: January 19, 2023



Peyman Adibi, Hassan Shahoon, Isfahan Gastroenterology and Hepatology Research Center, Isfahan University of Medical Sciences, Isfahan 8174673461, Iran

Simindokht Kalani, Department of Psychology, University of Isfahan, Isfahan 8174673441, Iran

Sayed Jalal Zahabi, Mohammad Reza Heidarpour, Department of Electrical and Computer Engineering, Isfahan University of Technology, Isfahan 8415683111, Iran

Homa Asadi, Mohammad Amouzadeh, Department of Linguistics, University of Isfahan, Isfahan 8174673441, Iran

Mohsen Bakhtiar, Department of Linguistics, Ferdowsi University of Mashhad, Mashhad 9177948974, Iran.

Hamidreza Roohafza, Department of Psychocardiology, Cardiac Rehabilitation Research Center, Cardiovascular Research Institute (WHO-Collaborating Center), Isfahan University of Medical Sciences, Isfahan 8187698191, Iran

Mohammad Amouzadeh, School of International Studies, Sun Yat-sen University, Zhuhai 519082, Guangdong Province, China

Corresponding author: Sayed Jalal Zahabi, PhD, Assistant Professor, Department of Electrical and Computer Engineering, Isfahan University of Technology, Isfahan 8415683111, Iran. zahabi@iut.ac.ir

Abstract

An important factor in the course of daily medical diagnosis and treatment is understanding patients' emotional states by the caregiver physicians. However, patients usually avoid speaking out their emotions when expressing their somatic symptoms and complaints to their non-psychiatrist doctor. On the other hand, clinicians usually lack the required expertise (or time) and have a deficit in mining various verbal and non-verbal emotional signals of the patients. As a result, in many cases, there is an emotion recognition barrier between the clinician and the patients making all patients seem the same except for their different somatic symptoms. In particular, we aim to identify and combine three major disciplines (psychology, linguistics, and data science) approaches for detecting emotions from verbal communication and propose an integrated solution for emotion recognition support. Such a platform may give emotional guides and indices to

the clinician based on verbal communication at the consultation time.

Key Words: Physician-Patient relations; Emotions; Verbal behavior; Linguistics; Psychology; Data science

©The Author(s) 2023. Published by Baishideng Publishing Group Inc. All rights reserved.

Core Tip: In the context of doctor-patient interactions, we focus on patient speech emotion recognition as a multifaceted problem viewed from three main perspectives: Psychology/psychiatry, linguistics, and data science. Reviewing the key elements and approaches within each of these perspectives, and surveying the current literature on them, we recognize the lack of a systematic comprehensive collaboration among the three disciplines. Thus, motivated by the necessity of such multidisciplinary collaboration, we propose an integrated platform for patient emotion recognition, as a collaborative framework towards clinical decision support.

Citation: Adibi P, Kalani S, Zahabi SJ, Asadi H, Bakhtiar M, Heidarpour MR, Roohafza H, Shahoon H, Amouzadeh M. Emotion recognition support system: Where physicians and psychiatrists meet linguists and data engineers. *World J Psychiatry* 2023; 13(1): 1-14

URL: <https://www.wjgnet.com/2220-3206/full/v13/i1/1.htm>

DOI: <https://dx.doi.org/10.5498/wjp.v13.i1.1>

INTRODUCTION

In order to establish a therapeutic relationship between physician and patient, it is necessary to have knowledgeable practitioners in various specialties as well as an effective interaction and communication between physician and patient which starts with obtaining the patient's medical history and continues to convey a treatment plan[1,2]. Doctor-patient communication is a complex interpersonal interaction where different types of expertise and techniques are required to understand this relationship completely in verbal and nonverbal forms, especially when trying to extract emotional states and determinants during a medical consultation session[3]. Doctor-patient communication is a complex interpersonal interaction which requires an understanding of each party's emotional state. In this paper, our focus is on physicians' understanding of patients' emotions. When patients attend medical consultation, they generally convey their particular experiences of the perceived symptoms to physicians. They interpret these somatic sensations in terms of many different factors including their unique personal and contextual circumstances. Motivated by the illness experience, they generate their own ideas and concerns (emotions), leading them to seek out consultation[4-6]. Generally, patients expect and value their doctors caring for these personal aspects of their experience[7,8]. During interactions and conversations with patients, physicians should be able to interpret their emotional states, which can help build up trust between patients and them[9,10]. This will ultimately lead to better clinical outcomes. Also, identifying and recording these states will help complete patients' medical records. Many diseases that seem to have physical symptoms are, in fact, largely intertwined with psychological variables, such as functional somatic syndromes (FSS)[11]. Increasingly, physicians have realized that recognizing the psychological state of patients with FSS will be very effective in providing an appropriate treatment. For example, the ability to accurately understand sound states may help interpret a patient's pain. Thus, the presence of information about patients' mental states in their medical records is essential.

Emotion detection accuracy, *i.e.*, the ability to detect whether a patient is expressing an emotion cue, has consequences for the physician-patient relationship. The key to patient-centered care is the ability to detect, accurately identify, and respond appropriately to the patient's emotions[12-15]. Failure to detect a patient's emotional cues may give rise to an ineffective interaction between doctor and patient, which may, in turn, lead to misdiagnosis, lower recall, mistreatments, and poorer health outcomes[16,17]. Indeed, if the emotion cue is never detected, then the ability to accurately identify or respond to the emotion never comes into play. Doctors who are more aware of their patients' emotions are more successful in treating them[13]. Patients have also reported greater satisfaction with such physicians[18-22]. Recognizing the emotions and feelings of patients provides the ground for more physician empathy with patients[23,24]. The academic and medical literature highlights the positive effects of empathy on patient care[25]. In this regard, the medical profession requires doctors to be both clinically competent and empathetic toward the patients. However, in practice, meeting both needs may be difficult for physicians (especially inexperienced and unskilled ones)[26]. On the other hand, patients do not always overtly express these experiences, feelings, concerns, and ideas. Rather, they often communicate them indirectly through more or less subtle nonverbal or verbal "clues" which nevertheless contain

interesting clinical information which can be defined as "clinical or contextual clues"[27-29]. They do not say, "Hey doctor, I'm feeling really emotional right now; or do you know whether I'm angry or sad?" Thus, emotional cues are often ambiguous and subtle[30-33].

On the other hand, patients' emotional audiences (*i.e.*, physicians) are often inexperienced in detecting emotions. One of the most important problems physicians face in the development of this process is the difficulty of capturing the clues that patients offer and failing to encourage them to expose details about these feelings[34]. Research indicates that over 70% of patients' emotional cues are missed by physicians [34]. It is unclear whether missed responses were the result from physicians detecting an emotional cue and choosing not to respond, or from failing to detect the cue in the first place. Indeed, these emotional cues present a challenge to doctors who often overlook them, as clinical information and therefore opportunities to know the patient's world are lost[34-37]. Physicians vary in their ability to recognize patients' emotions, with some being fully aware of the significance of understanding emotions and capable of identifying them. They also range from high emotional intelligence to low emotional intelligence. Another argument often heard from physicians is that they do not have time for empathy [38].

Despite the importance of such issues, this aspect remains grossly overlooked in conventional medical training. This comes from the fact that training emotion skills in medical schools is variable, lacks a strong evidence- base, and often does not include the training of emotion processing[39].

In the preceding paragraphs, four reasons were offered as to why physicians have failed to detect and interpret patients' emotional states, and hence why we need to find a solution for this problem. These reasons could be summarized as follows. First, detecting patients' emotions can contribute to healing them, as well as to increasing their satisfaction. Secondly, emotional cues are mostly indirectly found in patients' speech. That is, emotional cues can be very subtle and ambiguous. Further, many physicians do not possess enough experience to detect patients' emotions or even when they are skilled and experienced enough to do so, they do not have time to deal with it. In addition, training doctors to detect patients' emotions has been thoroughly overlooked in routine medical training. Thus, if a solution can be found to help physicians recognize patients' emotions and psychological states, this problem can be overcome to a large extent.

One strategy is to develop and employ a technology that can provide information about the patient's emotions, feelings, and mental states by processing their verbal and non-verbal indicators (Figure 1). In the present manuscript, we focus on verbal communication. Human speech carries a tremendous number of informative features, which enables listeners to extract a wealth of information about speakers' identity. These features can range from linguistic characteristics through extralinguistic features to paralinguistic information, such as the speaker's feelings, attitudes, or psychological states [40]. The psychological states (including emotions, feelings, and affections) embedded in people's speech are among the most important parts of the verbal communication array humans possess. As with other non-verbal cues, they are under conscious control much less than verbal cues. This makes speech an excellent guide to a human's "true" emotional state even when he/she is trying to hide it.

In order to design and present such technology, the first step is to know which indicators in speech can be used to identify emotions. Psychologists, psychiatrists, and linguists have done extensive research to identify people's emotions and feelings, and have identified a number of indicators. They believe that through these markers, people's emotions and feelings can be understood.

THE PSYCHOLOGICAL APPROACH

Psychologists and psychiatrists pay attention to content indicators and acoustic variables to identify people's emotions through their speech. Scholarly evidence suggests that mental health is associated with specific word use[41-43]. Psychologists and psychiatrists usually consider three types of word usage to identify emotions: (1) Positive and negative emotion words; (2) standard function word categories; and (3) content categories. They distinguish between positive ("happy", "laugh") and negative ("sad", "angry") emotion words, standard.

Function word categories (*e.g.*, self-references, first, second, and third person pronouns) and various content categories (*e.g.*, religion, death, and occupation). The frequent use of "You" and "I" suggests a different relationship between the speaker and the addressee than that of "We". The former suggests a more detached approach, whereas the latter expresses a feeling of solidarity. Multiple studies have indicated that the frequent use of the first-person singular is associated with negative affective states[44-48], which reveals a high degree of self-preoccupation[49]. People with negative emotional states (such as sadness or depression) use second and third person pronouns less often[38-40]. These people have a lower ability to express positive emotions and express more negative emotions in their speech[44-48]. Also, people with negative emotional states use more words referring to death[44].

In addition to the content of speech, psychologists and psychiatrists also look at several acoustic variables (such as pitch variety, pause time, speaking rate, and emphasis) to detect emotions. According to the research in this area, people with negative emotional states typically have a slower speaking rate [50-54], lower pitch variety[55,56], produce fewer words[57], and have longer pauses[53,54,58].

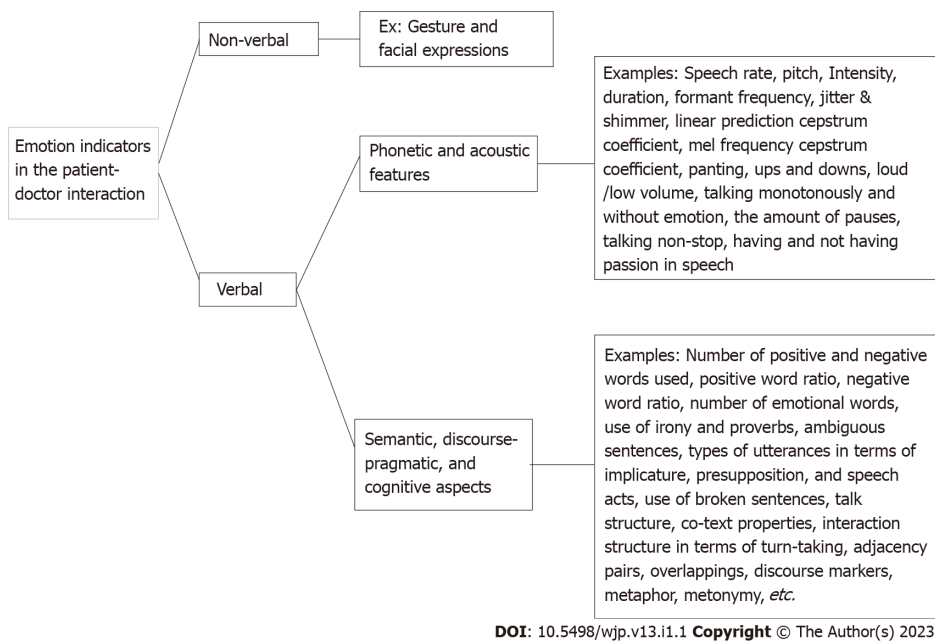


Figure 1 Emotion indicators in the patient-doctor interaction.

THE LINGUISTIC APPROACH

Within linguistics, various approaches (*e.g.*, phonetic, semantic, discourse-pragmatic, and cognitive) have been adopted to examine the relationship between language and emotion[56,59,60]. As far as the phonetic and acoustic studies are concerned, emotions can be expressed through speech and are typically accompanied with physiological signals such as muscle activity, blood circulation, heart rate, skin conductivity, and respiration. This will subsequently affect the kinematic properties of the articulators, which in turn will cause altered acoustic characteristics of the produced speech signals of the speakers. Studies of the effects of emotion on the acoustic characteristics of speech have revealed that parameters related to the frequency domain (*e.g.*, average values and ranges of fundamental frequency and formant frequencies), the intensity domain of speech (*e.g.*, energy, amplitude), temporal characteristics of speech (*e.g.*, duration and syllable rate), spectral features Mel frequency cepstral coefficients, and voice quality features (*e.g.*, jitter, shimmer, and harmonics-to-noise-ratio) are amongst the most important acoustically measurable parameters for correlates of emotion in speech. For instance, previous studies have reported that the mean and range of fundamental frequency observed for utterances spoken in anger situations were considerably greater than the mean and range for the neutral ones, while the average fundamental frequency for fear was lower than that observed for anger[61] (Figure 2 and Table 1).

Past research has produced many important findings to indicate that emotions can be distinguished by acoustical patterns; however, there are still a multitude of challenges regarding emotional speech research. One of the major obstacles that must be tackled in the domain of emotion recognition relates to variable vocalization which exists within speakers. Voices are often more variable within the same speaker (within-speaker variability) than they are between different speakers and it is thus unclear how human listeners can recognize individual speakers' emotion from their speech despite the tremendous variability that individual voices reveal. Emotion is sensitive to a large degree of variation within a single speaker and is highly affected by factors such as gender, speakers, speaking styles, sentence structure in spoken language, culture, and environment. Thus, identifying what specific mechanisms motivate variability in acoustic properties of emotional speech and how we can overcome differences arising from individual properties remain major challenges ahead of the emotion recognition field.

With regard to investigations in the area of pragmatics (in its continental notion which encompasses discourse analysis, sociolinguistics, cognitive linguistics, and even semantics), we observe a flourishing trend in linguistics focusing on the emotion in language[59,62]. These studies have examined important issues related to referential and non-referential meanings of emotion. In semantics, the focus has been on defining emotional and sentimental words and expressions, collocations and frames of emotion[63, 64], field semantics[62], as well as lexical relations including semantic extensions. However, more pragmatic and discourse-oriented studies have looked at issues in terms of emotion and cultural identity[65,66]; information structure/packaging (*e.g.* topicalization and thematicization[67] and emotion, emotive particles and interjections[68-70], emotional implicatures, and emotional illocutionary acts, deixis, and indexicality (*e.g.* proximalization and distalization[71,72], conversational analysis and emotion (*e.g.* turn-taking and interruption)[73,74], *etc.*

Table 1 Acoustic differences related to prosody and spectral features of the word (sahar) produced by a Persian female speaker in neutral and anger situations

	Neutral	Angry
Prosody features		
Mean Fundamental frequency (F0)	200 Hz	225 Hz
Minimum of the fundamental frequency	194 Hz	223 Hz
Maximum of the fundamental frequency	213 Hz	238 Hz
Mean intensity	60 dB	78 dB
Spectral features		
First formant frequency (F1)	853 Hz	686 Hz
Second formant frequency (F2)	2055 Hz	1660 Hz
Third formant frequency (F3)	3148 Hz	2847 Hz
Fourth formant frequency (F4)	4245 Hz	3678 Hz

Cognitive linguists use other methods to recognize emotion in speech. The cognitive linguistic approach to emotion concepts is based on the assumption that conventionalized language used to talk about emotions is a significant tool in discovering the structure and content of emotion concepts[75]. They consider a degree of universality for emotional experience and hold that this partial universality arises from basic image schemas that emerge from fundamental bodily experiences[76-79]. In this regard, the cultural model of emotions is a joint product of (possibly universal) actual human physiology, metonymic conceptualization of actual human physiology, metaphor, and cultural context [77]. In this approach, metaphor and metonymy are used as conceptual tools to describe the content and structure of emotion concepts.

Conceptual metaphors create correspondences between two distinct domains. One of the domains is typically more physical or concrete than the other (which is thus more abstract)[76]. For example, in the Persian expression *gham dar delam âshiyâneh kardeh* 'sadness has nested in my heart', *gham* 'sadness' is metaphorically conceptualized as a bird and *del* 'heart/stomach' is conceived of as a nest. The metaphor focuses on the perpetuation of sadness. The benefit of metaphors in the study of emotions is that they can highlight and address various aspects of emotion concepts[75,76]. Metonymy involves a single domain, or concept. Its purpose is to provide mental access to a domain through a part of the same domain (or vice versa) or to a part of a domain through another part in the same domain[80]. Metonymies can express physiological and behavioral aspects of emotions[75]. For example, in *she was scarlet with rage*, the physiological response associated with anger, i.e., redness in face and neck area, metonymically stands for anger. Thus, cognitive linguistics can contribute to the identification of metaphorical and metonymical conceptualizations of emotions in large corpora.

Although speech provides substantial information about the emotional states of speakers, accurate detection of emotions may nevertheless not always be feasible due to challenges that pervade communicative events involving emotions. Variations at semantic, pragmatic, and social-cultural levels present challenges that may hinder accurately identifying emotions *via* linguistic cues. At the semantic level, one limitation seems to be imposed by the "indeterminacy of meaning", a universal property of meaning construction which refers to "situations in which a linguistic unit is underspecified due to its vagueness in meaning"[81]. For example, Persian expressions such as *ye juriam* or *ye hâliam* roughly meaning 'I feel strange or unknown' even in context may not explicitly denote the emotion(s) the speaker intends to convey, and hence underspecify the conceptualizations that are linguistically coded. The other limitation at the semantic level pertains to cross-individual variations in the linguistic categorization of emotions. Individuals differ as to how they linguistically label their emotional experiences. For example, the expression *tu delam qoqâst* 'there is turmoil in my heart' might refer to 'extreme sadness' for one person but might suggest an 'extreme sense of confusion' for another. Individuals also reveal varying degrees of competence in expressing emotions. This latter challenge concerns the use of emotion words, where social categories such as age, gender, ethnic background, education, social class, and profession could influence the ease and skill with which speakers speak of their emotions. Since emotions perform different social functions in different social groups[82], their use is expected to vary across social groups.

Language differences are yet another source of variation in the use and expression of emotions, which presents further challenges to the linguistic identification of emotions. Each language has its own specific words, syntactic structures, and modes of expressions to encode emotions. Further, emotions are linked with cultural models and reflect cultural norms as well as values[83]. Thus, emotion words cannot be taken as culture-free analytical tools or as universal categories for describing emotions[84].

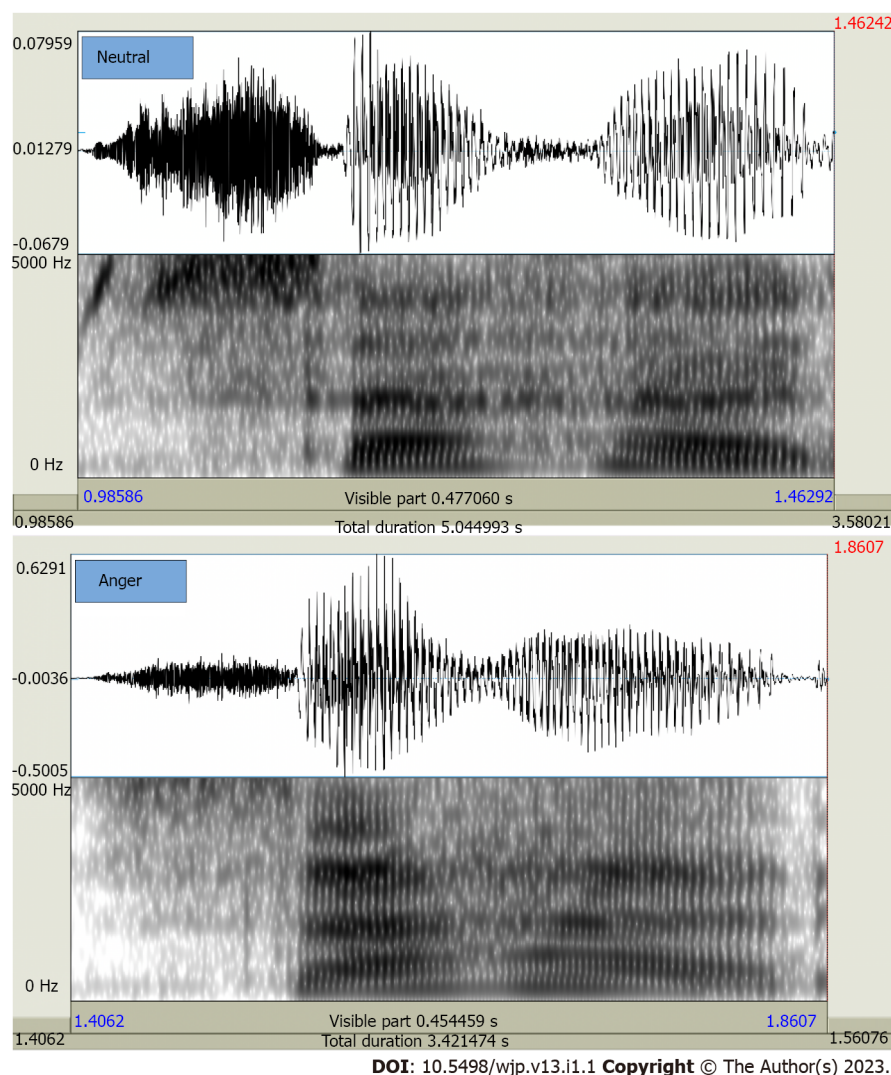


Figure 2 Spectrograms of the Persian word (sahar) pronounced by a Persian female speaker in neutral (top) and anger (down) situations.

Figure 2 shows spectrograms of the word (sahar), spoken by a native female speaker of Persian. The figure illustrates a couple of important differences between acoustic representations of the produced speech sounds. For example, the mean fundamental frequency in anger situations is higher (225 Hz) than that observed for neutral situations (200 Hz). Additionally, acoustic features such as mean formant frequencies (e.g. F1, F2, F3, and F4), minimum and maximum of the fundamental frequency, and mean intensity are lower in neutral situations. More details are provided in Table 1.

Patterns of communication vary across and within cultures. The link between communication and culture is provided by a set of shared interpretations which reflect beliefs, norms, values, and social practices of a relatively large group of people[85]. Cultural diversity may pose challenges to doctors and health care practitioners in the course of communicating with patients and detecting their emotions. In a health care setting, self-disclosure is seen as an important (culturally sensitive) characteristic that differentiates patients according to their degree of willingness to tell the doctor/practitioner what they feel, believe, or think[86]. Given the significance of self-disclosure and explicitness in the verbal expression of feelings in health care settings (Robinson, *ibid*), it could be predicted that patients coming from social groups with more indirect, more implicit, and emotionally self-restrained styles of communication will probably pose challenges to doctors in getting them to speak about their feelings in a detailed and accurate manner. In some ethnic groups, self-disclosure and intimate revelations of personal and social problems to strangers (people outside one's family or social group) may be unacceptable or taboo due to face considerations. Thus, patients belonging to these ethnic groups may adopt avoidance strategies in their communication with the doctor and hide or understate intense feelings. People may also refrain from talking about certain diseases or use circumlocutions due to the taboo or negative overtones associated with them. Further, self-restraint may be regarded as a moral virtue in some social groups, which could set a further obstacle in self-disclosing to the doctor or healthcare practitioner.

Overall, it is seen that these linguistically-oriented studies reveal important aspects of emotion in language use. In particular, they have shown how emotion is expressed and constructed by speakers in discourse. Such studies, however, are not based on multi-modal research to represent a comprehensive and unified description of emotion in language use. This means that, for a more rigorous and fine-

grained investigation, we need an integrative and cross-disciplinary approach to examining emotions in language use.

THE DATA SCIENCE APPROACH

From the data science perspective, speech emotion recognition (SER) is a machine learning (ML) problem whose goal is to classify the speech utterances based on their underlying emotions. This can be viewed from two perspectives: (1) Utterances as sounds with acoustic and spectral features (non-verbal); and (2) Utterances as words with specific semantic properties (verbal)[87-91]. While in the literature, SER typically refers to the former perspective, the latter is also important and provides a rich source of information, which can be harvested in favor of emotion recognition *via* natural language processing (NLP). Recent advances in the NLP technology allow for a fast analysis of text. In particular, word vector representations (also known as word embeddings) are used to embed words in a high dimensional space where words maintain semantic relationships with each other[92]. These vector representations, which are obtained through different ML algorithms, commonly capture the semantic relations between the words by looking into their collocation/co-occurrence in large corpora. In this way, the representation of each word and the machine's understanding of that partially reflect the essential knowledge that relates to that word, thus capturing the so-called frame semantics. The problem of SER can thus be tackled by analyzing the transcript of the speech by running various downstream tasks on the word vectors of the given speech.

As for the former perspective, different classifiers have so far been suggested for SER as candidates for a practically feasible automatic emotion recognition (AER) system. These classifiers can be put broadly into two main categories: Linear classifiers and non-linear classifiers. The main classification techniques/models within these two categories are: (1) Hidden Markov model[93-96]; (2) Gaussian mixture model[97,98]; (3) K-Nearest neighbor[99]; (4) Support vector machine[100,101]; (5) Artificial neural network[94,102]; (6) Bayes classifier[94]; (7) Linear discriminant analysis[103,104]; and (8) Deep neural network[102-107].

A review of the most relevant works within the above techniques has recently been done in[108]. We have provided a short description of the above techniques in Appendix. One of the main approaches in the last category, *i.e.*, deep neural networks, is to employ transfer learning. Recently[109] has reviewed the application of generalizable transfer learning in AER in the existing literature. In particular, it provides an overview of the previously proposed transfer learning methods for speech-based emotion recognition by listing 21 relevant studies.

The classifiers developed for SER may also be categorized in terms of their feature sets. Specifically, there are three main categories of speech features for SER: (1) The prosodic features[110-114]; (2) The excitation source features[110,111,115,116]; and (3) The spectral or vocal tract features[117-120].

Rosodic features, also known as continuous features, are some attributes of the speech sound such as pitch or fundamental frequency and energy. These features can be grouped into the following subcategories[104,105]: (1) Pitch-related features; (2) Formant features; (3) Energy-related features; (4) Timing features; and (5) Articulation features. Excitation source features, which are also referred to as voice quality features, are features which are used to represent glottal activity, such as harshness, breathiness, and tenseness of the speech signal.

Finally, spectral features, also known as segmental or system features, are the characteristics of various sound components generated from different cavities of the vocal tract system that have been extracted in different forms. The particular examples are ordinary linear predictor coefficients[117], one-sided autocorrelation linear predictor coefficients[113], short-time coherence method[114], and least squares modified Yule-Walker equations[115].

Table 2 summarizes the three discussed approaches to recognizing emotional indicators in speech 1.

Given the breadth and complexity of emotion detection indicators in psychology and linguistics, it is difficult to establish a decision support system for a doctor's emotional perception of patients. This requires a comprehensive and multidisciplinary approach. In order to build such a system, an application will be very useful. When a person experiences intense excitement, in addition to a reduction in his/her concentration, his/her mental balance is also disturbed more easily and quickly. This is also used as a strategy in sociology to take hold of people's minds.

Under unstable conditions, reasoning and logical thinking (and thus more effective and active behavior), which emerge in response to the activity of new and higher parts of the brain, are dominated by older parts of the brain, which have more biological precedents (several thousand *vs* millions of years). Thus, these older parts act impulsively or reactively.

Working in an emergency environment and sometimes even in an office has special conditions, such as excessive stress due to medical emergencies, pressure from patient companions, patient's own severe fear, as well as the impact of the phenomenon of "transference" and "countertransference" between physician and patient or between physician and patient companion. These can impair a physician's ability to reason and think logically. Thus, use of such an intelligent system can enhance doctors' efficiency, increase their awareness, and make it easier for them to manage the conditions.

Table 2 Different approaches to recognizing the emotional indicators in speech

Approaches	Emotional indicators
Psychological	(1) Positive and negative emotion words; (2) Standard function word categories; (3) Content categories; (4) The way of pronoun usage; and (5) Acoustic variables (such as pitch variety, pause time, speaking rate and emphasis)
Linguistic	(1) Phonetic: Spectral analysis, temporal analysis; (2) Semantic & Discourse-pragmatic: Words, field, cultural identity, emotional implicatures, illocutionary acts, deixis and indexicality; and (3) Cognitive: Metaphor, metonymy
Data science	(1) SER: Looking at sounds with acoustic and spectral features; and (2) NLP: Looking at words with specific semantic properties, word embedding

SER: Speech emotion recognition; NLP: Natural language processing.

THE PROPOSED SOLUTION

In the previous sections, the problem of SER was viewed from its three main perspectives: Psychology/psychiatry, linguistics, and data science, and the key elements within each perspective were highlighted. One way to integrate these three sides and benefit from their potential contributions to SER is through developing an intelligent platform. In what follows, focusing on SER in the context of doctor-patient interactions, we propose a solution for such integration.

The proposed solution consists of two key components: (1) The intelligent processing engine; and (2) The data-gathering platform.

The intelligent processing engine, at the algorithmic level, is based on NLP, speech processing, and in a wider context, behavioral signal processing methods. While it is clear that the processing engine will serve as the brain of the proposed intelligent platform, and is indeed a place where the novelty, creativity, and robustness of implemented algorithms can make a great difference, it will not practically function desirably without a well-thought, flexible data-gathering platform. Thus, despite the genuine algorithms which are to be developed at the core of the platform, and the undeniable impact they will have on the performance of the system, we believe it is the data-gathering platform that will make the solution very unique. One idea is to develop a cloud-based multi-mode multi-sided data gathering platform, which has three main sides: (1) The patient side; (2) The physician side; and (3) The linguistic/psychologist side.

Regarding the functioning of the platform, three modes can be considered: (1) The pre-visit mode; (2) The on-visit mode; and (3) The post-visit mode.

The pre-visit mode will include the patient's declaration of his/her health-related complaints/conditions and concerns, which will be automatically directed to the cloud-based processing engine, and labeled *via* a SER algorithm. This mode is reinforced *via* receiving additional multi-dimensional data from the patient through filling various forms and questionnaires. Also, it is possible for the patient to submit text to accompany his/her speech. This allows one to perform additional classification/clustering tasks such as sentiment analysis or patient segmentation on the provided text, using biomedical NLP methods. The on-visit mode enables the recording of the visiting session and the clinician-patient conversations. Finally, the post-visit mode of the application provides an interface for the psychiatrist/psychologist as well as the linguist to extract and label the psychological and linguistic features within the patient's speech. Such tagging of the data by a team of specialists will in the long term lead to a rich repository of patient speech, which is of great value in training the ML algorithms in the processing engine. The proposed platform, which we have named INDICES, is depicted in [Figure 3](#).

Although the proposed platform is to be designed such that it scales up at the population level in order to benefit from the diversity of the gathered data, it will also serve every individual as a customized personalized electronic health record that keeps track of the patient's psycho-emotional profile. As for the implementation of the platform, it is practically possible to tailor it to various devices (cell phones, tablets, PCs, and Laptops) *via* android/macOS, and web service applications

Note that emotion is essentially a multifaceted concept and no matter how sophisticated the proposed signal processing and data mining technology is, it would eventually face limitations in grasping all of its aspects. For instance, cultural aspects of expressing emotions can be a serious challenge to the technological system. Extracting the appropriate measurable features for correctly interpreting the cultural indices of emotion in speech can be a challenge, which nonetheless adds to the beauty of the problem. Further, as mentioned earlier, not all emotional indicators are embedded in the speech. Indeed, facial expressions and body gestures play important roles in expressing one's emotions as well. Hence, since the technology considered in our proposed method is focused merely on speech signals, it will of course have *blind spots* such as the visual aspects of emotion which are not exploited. This can be thought of as a main limitation that bounds the performance of the proposed emotion recognition system. However, with the same pattern that technology has always emerged throughout history, the proposed method can similarly serve as a baseline to which further improvements and additional capabilities can be added in future. We must also note that in capturing the different aspects of emotion, we are faced with

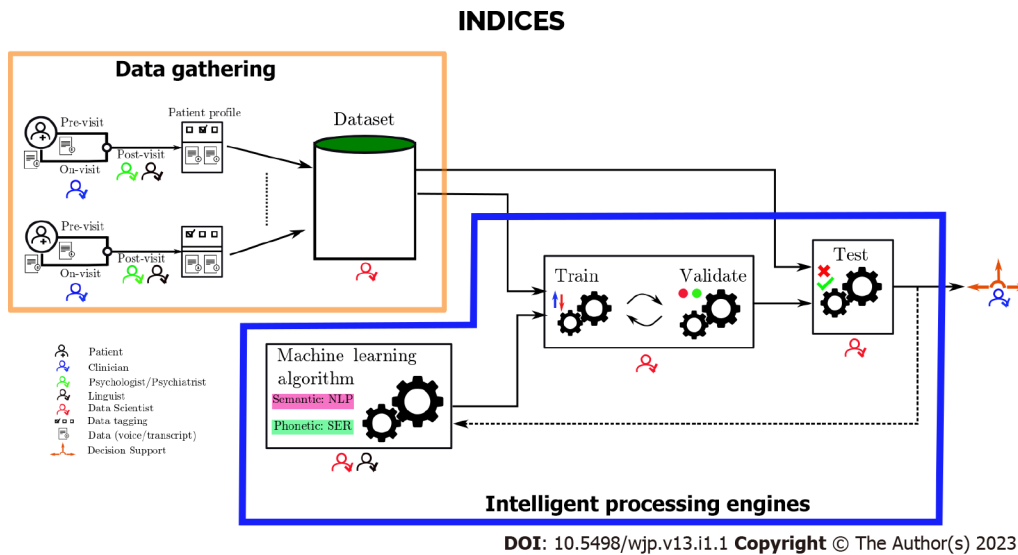


Figure 3 Integrated platform for patient emotion recognition and decision support. It consists of the data gathering platform and the intelligent processing engines. Each patient's data, in the form of voice/transcripts is captured, labeled, and stored in the dataset. The resulting dataset feeds the machine language training/validation and test engines. The entire process of intelligent processing may iterate several times for further fine tuning. It is crucial to have collaboration among the three relevant expertise in different parts of the proposed solution.

a tradeoff between computational complexity and performance. In particular, depending on the required accuracy of the system, one may need to customize the aspects of emotion which are to be examined *via* technology, taking into account the computational burden they would impose on the system.

We shall finally end this section with two remarks. First, it is important to note that despite all integrations and optimizations involved in the design and training of the proposed intelligent platform, it would still have the intrinsic limitations of a machine as a decision-maker, some of which were mentioned above. Thus, the proposed solution would eventually serve as a decision aid/support (and not as a decision replacement). Secondly, while the proposed solution provides a global framework, it invites for a series of methodologies and solutions, which are to be adapted and customized to each language and culture setting for local use.

APPENDIX

We provide Table 3, which includes a brief description of each of the data science techniques and models mentioned earlier, along with reference sources in which further technical details of the methods can be found.

CONCLUSION

In the context of doctor-patient interactions, this article focused on patient SER as a multidimensional problem viewed from three main aspects: Psychology/psychiatry, linguistics, and data science. We reviewed the key elements and approaches within each of these three perspectives, and surveyed the relevant literature on them. In particular, from the psychological/psychiatric perspective, the emotion indicators in the patient-doctor interaction were highlighted and discussed. In the linguistic approach, the relationship between language and emotion was discussed from phonetic, semantic, discourse-pragmatic, and cognitive perspectives. Finally, in the data science approach, SER was discussed as a ML/signal processing problem. The lack of a systematic comprehensive collaboration among the three discussed disciplines was pointed out. Motivated by the necessity of such multidisciplinary collaboration, we proposed a platform named indices: An integrated platform for patient emotion recognition and decision support. The proposed solution can serve as a collaborative framework towards clinical decision support.

Table 3 A brief description of some data science models/methods

Method/Model	Short description	Ref.
HMM	A HMM is a statistical model that can be used to describe the evolution of observable events that depend on internal factors, which are not directly observable. The observed event is called a 'symbol' and the invisible factor underlying the observation is called a 'state'. A HMM consists of two stochastic processes, namely, an invisible process of hidden states and a visible process of observable symbols. The hidden states form a Markov chain and the probability distribution of the observed symbol depend on the underlying state. Via this model, the observations are modeled in two layers: One visible and the other invisible. Thus, it is useful in classification problems where raw observations are to be put into a number of categories that are more meaningful to us (Supplementary Figure 1)	[121, 122]
Gaussian mixture model	A Gaussian mixture model is a probabilistic model that assumes all data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters (Supplementary Figure 2)	[123]
KNN	KNN is a type of supervised learning algorithm used for classification. KNN tries to predict the correct class for the test data by calculating the distance between the test data and all training points. The algorithm then selects the K number of points which are closest to the test data. The KNN algorithm calculates the probability of the test data belonging to the classes of 'K' training data where the class that holds the highest probability (by majority voting) will be selected (Supplementary Figure 3)	[123]
SVM	The SVM is an algorithm that finds a hyperplane in an N-dimensional space (N: The number of features) that distinctly classifies the data points in a way that the plane has the maximum margin, <i>i.e.</i> , the maximum distance between data points of the two classes. Maximizing this margin distance would allow the future test points to be classified more accurately. Support vectors are data points that are closer to the hyperplane and influence the position as well as orientation of the hyperplane (Supplementary Figure 4)	[123]
Artificial neural network	An artificial neural network is a network of interconnected artificial neurons. An artificial neuron which is inspired by the actual neuron is modeled with inputs which are multiplied by weights, and then passed to a mathematical function which determines the activation of the neuron. The neurons in a neural network are grouped into layers. There are three main types of layers: - Input Layer - Hidden Layer(s) - Output Layer. Depending on the architecture of the network, outputs of some neurons are carried along with certain weights as inputs to some other neurons. By passing an input through these layers, the neural network finally outputs a value (discrete or continuous) which can be used to perform various classification/regression tasks. In this context, the neural network first has to learn the set of weights <i>via</i> the patterns within the so called training dataset, which is a sufficiently large set of input data labeled with their corresponding correct (expected) output (Supplementary Figure 5)	[124]
Bayes classifier	Bayes classifier, which is based on Bayes' theorem in probability, models the probabilistic relationships between the feature set and the class variable. Based on the modeled relationships, it estimates the class membership probability of the unseen example, in such a way that it minimizes the probability of misclassification	[123]
Linear discriminant analysis	Linear discriminant analysis is a method used in statistical machine learning, to find a linear combination of features that characterizes or separates two or more classes of objects or events. The resulting linear combination can be used as a linear classifier, or, as a means to dimension reduction prior to the actual classification task	[124]

HMM: Hidden markov model; KNN: K-nearest neighbor; SVM: Support vector machine.

FOOTNOTES

Author contributions: Adibi P, Kalani S, Zahabi SJ, Asadi H, Bakhtiar M, Heidarpour MR, Roohafza H, Shahoon H, Amouzadeh M; all contributed in conceptualization, identifying relevant studies, framing the results; Kalani S and Roohafza H wrote the psychological related part of the paper; Zahabi SJ and Heidarpour MR wrote the data science related part of the paper; Asadi H, Bakhtiar M, and Amouzadeh M, wrote the phonetics-linguistic, cognitive-linguistic, and semantic-linguistic related parts of the paper, respectively; Adibi P, Roohafza H, Shahoon H supervised the study.

Conflict-of-interest statement: All the authors report no relevant conflicts of interest for this article.

Open-Access: This article is an open-access article that was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution NonCommercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <https://creativecommons.org/licenses/by-nc/4.0/>

Country/Territory of origin: Iran

ORCID number: Peyman Adibi 0000-0001-6411-5235; Simindokht Kalani 0000-0002-9999-541X; Sayed Jalal Zahabi 0000-0001-5868-8192; Homa Asadi 0000-0003-1655-1336; Mohsen Bakhtiar 0000-0001-7012-6619; Mohammad Reza Heidarpour 0000-0002-2819-2556; Hamidreza Roohafza 0000-0003-3582-0431; Hassan Shahoon 0000-0003-1945-3520; Mohammad Amouzadeh 0000-0001-8964-7967.

S-Editor: Liu XF

L-Editor: A

P-Editor: Liu XF

REFERENCES

- 1 **Riedl D**, Schüßler G. The Influence of Doctor-Patient Communication on Health Outcomes: A Systematic Review. *Z Psychosom Med Psychother* 2017; **63**: 131-150 [PMID: 28585507 DOI: 10.13109/zptm.2017.63.2.131]
- 2 **Begum T**. Doctor patient communication: A review. *J Bangladesh Coll Phys Surg* 2014; **32**: 84-88 [DOI: 10.3329/jbcps.v32i2.26036]
- 3 **Kee JWY**, Khoo HS, Lim I, Koh MYH. Communication Skills in Patient-Doctor Interactions: Learning from Patient Complaints. *Heal Prof Educ* 2018; **4**: 97-106 [DOI: 10.1016/j.hpe.2017.03.006]
- 4 **Helman CG**. Communication in primary care: The role of patient and practitioner explanatory models. *Soc Sci Med* 1985; **20**: 923-931 [DOI: 10.1016/0277-9536(85)90348-X]
- 5 **Kleinmann A**. The illness narratives. USA: Basic Books, 1988
- 6 **McWhinney IR**. Beyond diagnosis: An approach to the integration of behavioral science and clinical medicine. *N Engl J Med* 1972; **287**: 384-387 [PMID: 5043523 DOI: 10.1056/NEJM197208242870805]
- 7 **Colliver JA**, Willis MS, Robbs RS, Cohen DS, Swartz MH. Assessment of Empathy in a Standardized-Patient Examination. *Teach Learn Med* 1998; **10**: 8-11 [DOI: 10.1207/S15328015TLM1001_2]
- 8 **Mercer SW**, Maxwell M, Heaney D, Watt GC. The consultation and relational empathy (CARE) measure: Development and preliminary validation and reliability of an empathy-based consultation process measure. *Fam Pract* 2004; **21**: 699-705 [PMID: 15528286 DOI: 10.1093/fampra/cmh621]
- 9 **Kadadi S**, Bharamanaiker S. Role of emotional intelligence in healthcare industry. *Drishtikon Manag J* 2020; **11**: 37
- 10 **Weng HC**. Does the physician's emotional intelligence matter? *Health Care Manage Rev* 2008; **33**: 280-288 [PMID: 18815493 DOI: 10.1097/01.HCM.0000318765.52148.b3]
- 11 **Barsky AJ**, Borus JF. Functional somatic syndromes. *Ann Intern Med* 1999; **130**: 910-921 [PMID: 10375340 DOI: 10.7326/0003-4819-130-11-199906010-00016]
- 12 **Beach MC**, Inui T; Relationship-Centered Care Research Network. Relationship-centered care. A constructive reframing. *J Gen Intern Med* 2006; **21** Suppl 1: S3-S8 [PMID: 16405707 DOI: 10.1111/j.1525-1497.2006.00302.x]
- 13 **Blue AV**, Chessman AW, Gilbert GE, Mainous AG 3rd. Responding to patients' emotions: Important for standardized patient satisfaction. *Fam Med* 2000; **32**: 326-330 [PMID: 10820674]
- 14 **Finset A**. "I am worried, Doctor! *Patient Educ Couns* 2012; **88**: 359-363 [PMID: 22819270 DOI: 10.1016/j.pec.2012.06.022]
- 15 **Mead N**, Bower P. Patient-centredness: A conceptual framework and review of the empirical literature. *Soc Sci Med* 2000; **51**: 1087-1110 [PMID: 11005395 DOI: 10.1016/S0277-9536(00)00098-8]
- 16 **Zimmermann C**, Del Piccolo L, Finset A. Cues and concerns by patients in medical consultations: A literature review. *Psychol Bull* 2007; **133**: 438-463 [PMID: 17469986 DOI: 10.1037/0033-2909.133.3.438]
- 17 **Jansen J**, van Weert JC, de Groot J, van Dulmen S, Heeren TJ, Bensing JM. Emotional and informational patient cues: The impact of nurses' responses on recall. *Patient Educ Couns* 2010; **79**: 218-224 [PMID: 20005066 DOI: 10.1016/j.pec.2009.10.010]
- 18 **Weng HC**, Chen HC, Chen HJ, Lu K, Hung SY. Doctors' emotional intelligence and the patient-doctor relationship. *Med Educ* 2008; **42**: 703-711 [PMID: 18588649 DOI: 10.1111/j.1365-2923.2008.03039.x]
- 19 **Hall JA**, Roter DL, Blanch DC, Frankel RM. Nonverbal sensitivity in medical students: Implications for clinical interactions. *J Gen Intern Med* 2009; **24**: 1217-1222 [PMID: 19771481 DOI: 10.1007/s11606-009-1107-5]
- 20 **DiMatteo MR**, Hays RD, Prince LM. Relationship of physicians' nonverbal communication skill to patient satisfaction, appointment noncompliance, and physician workload. *Health Psychol* 1986; **5**: 581-594 [PMID: 3803351 DOI: 10.1037//0278-6133.5.6.581]
- 21 **DiMatteo MR**, Taranta A, Friedman HS, Prince LM. Predicting patient satisfaction from physicians' nonverbal communication skills. *Med Care* 1980; **18**: 376-387 [PMID: 7401698 DOI: 10.1097/00005650-198004000-00003]
- 22 **Kim SS**, Kaplowitz S, Johnston MV. The effects of physician empathy on patient satisfaction and compliance. *Eval Health Prof* 2004; **27**: 237-251 [PMID: 15312283 DOI: 10.1177/0163278704267037]
- 23 **Shi M**, Du T. Associations of emotional intelligence and gratitude with empathy in medical students. *BMC Med Educ* 2020; **20**: 116 [PMID: 32303212 DOI: 10.1186/s12909-020-02041-4]
- 24 **Arora S**, Ashrafian H, Davis R, Athanasiou T, Darzi A, Sevdalis N. Emotional intelligence in medicine: A systematic review through the context of the ACGME competencies. *Med Educ* 2010; **44**: 749-764 [PMID: 20633215 DOI: 10.1111/j.1365-2923.2010.03709.x]
- 25 **Hojat M**, Louis DZ, Maio V, Gonnella JS. Empathy and health care quality. *Am J Med Qual* 2013; **28**: 6-7 [PMID: 23288854 DOI: 10.1177/1062860612464731]
- 26 **Ogle J**, Bushnell JA, Caputi P. Empathy is related to clinical competence in medical care. *Med Educ* 2013; **47**: 824-831 [PMID: 23837429 DOI: 10.1111/medu.12232]
- 27 **Marvel MK**. Involvement with the psychosocial concerns of patients. Observations of practicing family physicians on a university faculty. *Arch Fam Med* 1993; **2**: 629-633 [PMID: 8118582 DOI: 10.1001/archfami.2.6.629]
- 28 **Byrne PS**, Long BE. Doctors Talking to Patients. London: National government publication, 1976
- 29 **Thompson BM**, Teal CR, Scott SM, Manning SN, Greenfield E, Shada R, Haidet P. Following the clues: Teaching medical students to explore patients' contexts. *Patient Educ Couns* 2010; **80**: 345-350 [PMID: 20674240 DOI: 10.1016/j.pec.2010.06.035]
- 30 **Zimmermann C**, Del Piccolo L, Bensing J, Bergvik S, De Haes H, Eide H, Fletcher I, Goss C, Heaven C, Humphris G, Kim YM, Langewitz W, Meeuwesen L, Nuebling M, Rimondini M, Salmon P, van Dulmen S, Wissow L, Zandbelt L, Finset A. Coding patient emotional cues and concerns in medical consultations: The Verona coding definitions of emotional sequences (VR-CoDES). *Patient Educ Couns* 2011; **82**: 141-148 [PMID: 20430562 DOI: 10.1016/j.pec.2010.03.017]
- 31 **Mjaaland TA**, Finset A, Jensen BF, Gulbrandsen P. Patients' negative emotional cues and concerns in hospital consultations: A video-based observational study. *Patient Educ Couns* 2011; **85**: 356-362 [PMID: 21392928 DOI: 10.1016/j.pec.2011.03.017]

- 10.1016/j.pec.2010.12.031]
- 32 **Del Piccolo L**, Goss C, Bergvik S. The fourth meeting of the Verona Network on Sequence Analysis "Consensus finding on the appropriateness of provider responses to patient cues and concerns". *Patient Educ Couns* 2006; **61**: 473-475 [DOI: 10.1016/S0738-3991(06)00132-7]
- 33 **Piccolo LD**, Goss C, Zimmermann C. The Third Meeting of the Verona Network on Sequence Analysis. Finding common grounds in defining patient cues and concerns and the appropriateness of provider responses. *Patient Educ Couns* 2005; **57**: 241-244 [PMID: 15911200 DOI: 10.1016/j.pec.2005.03.003]
- 34 **Levinson W**, Gorawara-Bhat R, Lamb J. A study of patient clues and physician responses in primary care and surgical settings. *JAMA* 2000; **284**: 1021-1027 [PMID: 10944650 DOI: 10.1001/jama.284.8.1021]
- 35 **Branch WT**, Malik TK. Using 'windows of opportunities' in brief interviews to understand patients' concerns. *JAMA* 1993; **269**: 1667-1668 [PMID: 8455300 DOI: 10.1001/jama.1993.03500130081036]
- 36 **Bylund CL**, Makoul G. Examining empathy in medical encounters: an observational study using the empathic communication coding system. *Health Commun* 2005; **18**: 123-140 [PMID: 16083407 DOI: 10.1207/s15327027hc1802_2]
- 37 **Easter DW**, Beach W. Competent patient care is dependent upon attending to empathic opportunities presented during interview sessions. *Curr Surg* 2004; **61**: 313-318 [PMID: 15165773 DOI: 10.1016/j.cursur.2003.12.006]
- 38 **Mjaaland TA**, Finset A, Jensen BF, Gulbrandsen P. Physicians' responses to patients' expressions of negative emotions in hospital consultations: A video-based observational study. *Patient Educ Couns* 2011; **84**: 332-337 [PMID: 21454033 DOI: 10.1016/j.pec.2011.02.001]
- 39 **Satterfield JM**, Hughes E. Emotion skills training for medical students: a systematic review. *Med Educ* 2007; **41**: 935-941 [PMID: 17822414 DOI: 10.1111/j.1365-2923.2007.02835.x]
- 40 **Rose P**. Forensic speaker identification. New York: Taylor & Francis, 2001
- 41 **Gottschalk LA**, Gleser, GC. The measurement of psychological states through the content analysis of verbal behavior. California: University of California Press, 1979
- 42 **Rosenberg SD**, Tucker GJ. Verbal behavior and schizophrenia. The semantic dimension. *Arch Gen Psychiatry* 1979; **36**: 1331-1337 [PMID: 496553 DOI: 10.1001/archpsyc.1979.01780120061008]
- 43 **Stiles WB**. Describing talk: A taxonomy of verbal response modes. *Lang Soc* 1993; **22**: 568-570 [DOI: 10.1017/S0047404500017504]
- 44 **Pennebaker JW**, Francis, ME, & Booth, RJ Linguistic inquiry and word count: LIWC 2001. Mahway: Lawrence Erlbaum Associates, 2001
- 45 **Weintraub W**. Verbal Behavior in Everyday Life. New York: Springer, 1989
- 46 **Bucci W**, Freedman N. The language of depression. *Bull Menninger Clin* 1981; **45**: 334-358 [PMID: 6176285]
- 47 **Weintraub W**. Verbal behavior: Adaptation and psychopathology. New York: Springer Publishing Company, 1981
- 48 **Rude SS**, Gortner E-M, Pennebaker JW. Language use of depressed and depression-vulnerable college students. *Cogn Emot*. 2004; **18**: 1121-1133 [DOI: 10.1080/02699930441000030]
- 49 **Balsters MJH**, Krahmer EJ, Swerts MG, Vingerhoets AJJM. Verbal and nonverbal correlates for depression: A review. *Curr Psychiatry Rev* 2012; **8**: 227-234 [DOI: 10.2174/157340012800792966]
- 50 **Kraepelin E**. Manic-depressive insanity and paranoia. Edinburgh UK: Alpha Editions, 1921
- 51 **Newman S**, Mather VG. Analysis of spoken language of patients with affective disorders. *Am J Psychiatry* 1938; **94**: 913-942 [DOI: 10.1176/ajp.94.4.913]
- 52 **Hinchliffe MK**, Lancashire M, Roberts FJ. Depression: Defence mechanisms in speech. *Br J Psychiatry* 1971; **118**: 471-472 [PMID: 5576651 DOI: 10.1192/bjp.118.545.471]
- 53 **Mundt JC**, Snyder PJ, Cannizzaro MS, Chappie K, Geralt DS. Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. *J Neurolinguistics* 2007; **20**: 50-64 [PMID: 21253440 DOI: 10.1016/j.jneuroling.2006.04.001]
- 54 **Sobin C**, Alpert M. Emotion in speech: The acoustic attributes of fear, anger, sadness, and joy. *J Psycholinguist Res* 1999; **28**: 347-365 [PMID: 10380660 DOI: 10.1023/a:1023237014909]
- 55 **Nilsson A**. Acoustic analysis of speech variables during depression and after improvement. *Acta Psychiatr Scand* 1987; **76**: 235-245 [PMID: 3673650 DOI: 10.1111/j.1600-0447.1987.tb02891.x]
- 56 **Alpert M**, Pouget ER, Silva RR. Reflections of depression in acoustic measures of the patient's speech. *J Affect Disord* 2001; **66**: 59-69 [PMID: 11532533 DOI: 10.1016/s0165-0327(00)00335-9]
- 57 **Weintraub W**, Aronson H. The application of verbal behavior analysis to the study of psychological defense mechanisms. IV. Speech pattern associated with depressive behavior. *J Nerv Ment Dis* 1967; **144**: 22-28 [PMID: 6023378 DOI: 10.1097/00005053-196701000-00005]
- 58 **Chapple ED**, Lindemann E. Clinical Implications of Measurements of Interaction Rates in Psychiatric Interviews. *Appl Anthropol* 1942; **1**: 1-11 [DOI: 10.17730/HUMO.1.2.3R61TWT687148GIJ]
- 59 **Prakash M**, Language and Cognitive Structures of Emotion. Cambridge: Palgrave Macmillan, 2016: 182
- 60 **Dresner E**, Herring SC. Functions of the nonverbal in CMC: Emoticons and illocutionary force. *Communication Theory* 2010; **20**: 249-268 [DOI: 10.1111/j.1468-2885.2010.01362.x]
- 61 **Williams CE**, Stevens KN. Emotions and speech: Some acoustical correlates. *J Acoust Soc Am* 1972; **52**: 1238-1250 [PMID: 4638039 DOI: 10.1121/1.1913238]
- 62 **Liu Y**. The emotional geographies of language teaching. *Teacher Development* 2016; **20**: 482-497 [DOI: 10.1080/13664530.2016.1161660]
- 63 **Ruppenhofer J**. The treatment of emotion vocabulary in FrameNet: Past, present and future developments. Düsseldorf University Press, 2018
- 64 **Johnson-Laird PN**, Oatley K. Emotions, music, and literature in: Ewis M, Haviland-Jones JM, Barrett LF: Handbook of emotions. London: Guilford Press, 2008: 102-113
- 65 **Giorgi K**. Emotions, Language and Identity on the Margins of Europe. London: Springer, 2014

- 66 Wilce JM, Wilce JM. Language and emotion. Cambridge: Cambridge University Press, 2009
- 67 Wang L, Bastiaansen M, Yang Y, Hagoort P. ERP evidence on the interaction between information structure and emotional salience of words. *Cogn Affect Behav Neurosci* 2013; **13**: 297-310 [PMID: 23292996 DOI: 10.3758/s13415-012-0146-2]
- 68 Braber N. Emotional and emotive language: Modal particles and tags in unified Berlin. *J Pragmat* **38**: 1487-503 [DOI: 10.1016/j.pragma.2005.10.003]
- 69 Alba-Juez L, Larina TV. Language and emotion: Discourse-pragmatic perspectives. *Russ J Linguist* 2018; **22**: 9-37 [DOI: 10.22363/2312-9182-2018-22-1-9-37]
- 70 Goddard C. Interjections and emotion (with special reference to "surprise" and "disgust"). *Emotion Review* 2014; **6**: 53-63 [DOI: 10.1177/1754073913491843]
- 71 Glazer T. The Semiotics of Emotional Expression. *Trans Charles S Peirce Soc* 2017; **53**: 189-215
- 72 Wilce JM. Current emotion research in linguistic anthropology. *Emot Rev* 2014; **6**: 77-85
- 73 Peräkylä A, Sorjonen ML. Emotion in interaction. New York: Oxford University Press, 2012
- 74 Stevanovic M, Peräkylä A. Experience sharing, emotional reciprocity, and turn-taking. *Front Psychol* 2015; **6**: 450 [PMID: 25926811 DOI: 10.3389/fpsyg.2015.00450]
- 75 Kövecses Z. Emotion concepts. New York: Springer, 1990
- 76 Kövecses Z. Metaphors of anger, pride and love. Amsterdam: Benjamins, 1986
- 77 Kövecses Z. Metaphor and emotion: Language, culture, and body in human feeling. Cambridge: Cambridge University Press, 2003
- 78 Lakoff G, Kövecses Z. The cognitive model of anger inherent in American English. In Holland D, Quinn N, Editors: Cultural models in language and thought. Cambridge: Cambridge University Press, 1987: 195-221
- 79 Yu N. The contemporary theory of metaphor: A perspective from Chinese. Amsterdam: John Benjamins Publishing, 1998
- 80 Kövecses Z, Radden G. Metonymy: Developing a cognitive linguistic view. *Cogn Linguist* 1998; **9**: 37-78
- 81 Radden G, Köpcke KM, Berg T, Siemund P. The construction of meaning in language. Aspects of Meaning Construction. Amsterdam: John Benjamins Publishing Co, 2007: 1-5
- 82 Salmela M. The functions of collective emotions in social groups. In Institutions, emotions, and group agents. Dordrecht: Springer, 2014: 159-176
- 83 Kövecses Z. The concept of emotion: Further metaphors. In: Emotion concepts. New York: Springer, 1990: 160-181
- 84 Wierzbicka A. Talking about emotions: Semantics, culture, and cognition. In: *Cognition & Emotion* 1992; **6**: 285-319
- 85 Lustig M, Koester J. Intercultural communication: Interpersonal communication across cultures. J. Koester-Boston: Pearson Education, 2010
- 86 Robinson NM. To tell or not to tell: Factors in self-disclosing mental illness in our everyday relationships (Doctoral dissertation). Available from: https://mars.gmu.edu/jspui/bitstream/handle/1920/7872/Robinson_dissertation_2012.pdf
- 87 Akçay MB, Oğuz K. Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers. *Speech Commun* 2020; **116**: 56-76 [DOI: 10.1016/j.specom.2019.12.001]
- 88 Tan L, Yu K, Lin L, Cheng X, Srivastava G, Lin JC, Wei W. Speech Emotion Recognition Enhanced Traffic Efficiency Solution for Autonomous Vehicles in a 5G-Enabled Space-Air-Ground Integrated Intelligent Transportation System. *IEEE T Intell Transp* 2022; **23**: 2830-42 [DOI: 10.1109/tits.2021.3119921]
- 89 Schuller BW. Speech emotion recognition: Two decades in a nutshell, benchmarks, and ongoing trends. *COMMUN ACM* 2018; **61**: 90-9 [DOI: 10.1145/3129340]
- 90 Zhang S, Zhang S, Huang T, Gao W. Speech Emotion Recognition Using Deep Convolutional Neural Network and Discriminant Temporal Pyramid Matching. *IEEE Trans Multimedia* 2018; **20**: 1576-1590 [DOI: 10.1109/tmm.2017.2766843]
- 91 Chen M, He X, Yang J, Zhang H. 3-D Convolutional Recurrent Neural Networks with Attention Model for Speech Emotion Recognition. *IEEE Signal Process Lett* 2018; **25**: 1440-1444 [DOI: 10.1109/Lsp.2018.2860246]
- 92 Samadi MA, Akhondzadeh MS, Zahabi SJ, Manshaei MH, Maleki Z, Adibi P. Evaluating Sparse Interpretable Word Embeddings for Biomedical Domain. 2020 Preprint. Available from: <https://arxiv.org/abs/2005.05114> [DOI: 10.48550/arXiv.2005.05114]
- 93 Bitouk D, Verma R, Nenkov A. Class-Level Spectral Features for Emotion Recognition. *Speech Commun* 2010; **52**: 613-625 [PMID: 23794771 DOI: 10.1016/j.specom.2010.02.010]
- 94 Fernandez R, Picard R. Modeling drivers' speech under stress. *Speech Commun* 2003; **40**: 145-159 [DOI: 10.1016/S0167-6393(02)00080-8]
- 95 Nwe T, Foo S, De Silva L. Speech emotion recognition using hidden Markov models. *Speech Commun* 2003; **41**: 603-623 [DOI: 10.1016/S0167-6393(03)00099-2]
- 96 Lee C, Yildirim S, Bulut M, Busso C, Kazemzadeh A, Lee S, Narayanan S. Effects of emotion on different phoneme classes. *J Acoust Soc Am* 2004; **116**: 2481-2481 [DOI: 10.1121/1.4784911]
- 97 Breazeal C, Aryananda L. Recognition of Affective Communicative Intent in Robot-Directed Speech. *Auton Robots* 2002; **12**: 83-104 [DOI: 10.1023/A:1013215010749]
- 98 Slaney M, McRoberts G. BabyEars: A recognition system for affective vocalizations. *Speech Commun* 2003; **39**: 367-384 [DOI: 10.1016/S0167-6393(02)00049-3]
- 99 Pao TL, Chen YT, Yeh JH, Liao WY. Combining acoustic features for improved emotion recognition in mandarin speech. In: Tao J, Tan T, Picard RW, editors. Affective Computing and Intelligent Interaction. International Conference on Affective Computing and Intelligent Interaction; 2005 Oct; Berlin. Heidelberg: Springer, 2005: 279-285
- 100 Wu S, Falk T, Chan W. Automatic speech emotion recognition using modulation spectral features. *Speech Commun* 2011; **53**: 768-785 [DOI: 10.1016/j.specom.2010.08.013]
- 101 Pierre-Yves O. The production and recognition of emotions in speech: Features and algorithms. *Int J Hum Comput* 2003; **1**: 157-183 [DOI: 10.1016/S1071-5819(02)00141-6]

- 102 **Zhu A**, Luo Q. Study on speech emotion recognition system in e-learning. In: Jacko A editors. Human-Computer Interaction. HCI Intelligent Multimodal Interaction Environments. International Conference on Human-Computer Interaction; 2007 Jul 22, Berlin, Heidelberg: Springer, 2007: 544-552
- 103 **Chen L**, Mao X, Xue Y, Cheng LL. Speech emotion recognition: Features and classification models. *Digital Signal Processing* 2012; **22**: 1154-160 [DOI: [10.1016/j.dsp.2012.05.007](https://doi.org/10.1016/j.dsp.2012.05.007)]
- 104 **Xanthopoulos P**, Pardalos PM, Trafalis TB. Linear discriminant analysis. In Xanthopoulos P, Pardalos PM, Trafalis TB. Robust data mining New York: Springer, 2013: 27-33
- 105 **Chen M**, He X, Yang J, Zhang H. 3-D Convolutional Recurrent Neural Networks with Attention Model for Speech Emotion Recognition. *IEEE Signal Process Let* 2018; **25**: 1440-1444 [DOI: [10.1109/LSP.2018.2860246](https://doi.org/10.1109/LSP.2018.2860246)]
- 106 **Zhang S**, Zhang S, Huang T, Gao W. Speech Emotion Recognition Using Deep Convolutional Neural Network and Discriminant Temporal Pyramid Matching. *IEEE Trans Multimedia* 2018; **20**: 1576-1590 [DOI: [10.1109/TMM.2017.2766843](https://doi.org/10.1109/TMM.2017.2766843)]
- 107 **Mao Q**, Dong M, Huang Z, Zhan Y. Learning salient features for speech emotion recognition using convolutional neural networks. *IEEE Trans Multimedia* 2014; **16**: 2203-2213 [DOI: [10.1109/TMM.2014.2360798](https://doi.org/10.1109/TMM.2014.2360798)]
- 108 **Feng K**, Chaspari T. A Review of Generalizable Transfer Learning in Automatic Emotion Recognition. *Front Comput Sci* 2020; **2** [DOI: [10.3389/fcomp.2020.00009](https://doi.org/10.3389/fcomp.2020.00009)]
- 109 **Roy T**, Marwala T, Chakraverty S. A survey of classification techniques in speech emotion recognition. In: Chakraverty S: Mathematical Methods in Interdisciplinary Sciences. New Jersey: Wiley, 2020: 33-48
- 110 **Cowie R**, Douglas-Cowie E, Tsapatsoulis N, Votsis G, Kollias S, Fellenz W, Taylor JG. Emotion recognition in human-computer interaction. *IEEE Signal Process Mag* 2001; **18**: 32-80 [DOI: [10.1109/79.911197](https://doi.org/10.1109/79.911197)]
- 111 **Murray IR**, Arnott JL. Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *J Acoust Soc Am* 1993; **93**: 1097-1108 [PMID: [8445120](https://pubmed.ncbi.nlm.nih.gov/8445120/) DOI: [10.1121/1.405558](https://doi.org/10.1121/1.405558)]
- 112 **Banase R**, Scherer KR. Acoustic profiles in vocal emotion expression. *J Pers Soc Psychol* 1996; **70**: 614-636 [PMID: [8851745](https://pubmed.ncbi.nlm.nih.gov/8851745/) DOI: [10.1037/0022-3514.70.3.614](https://doi.org/10.1037/0022-3514.70.3.614)]
- 113 **Beeke S**, Wilkinson R, Maxim J. Prosody as a compensatory strategy in the conversations of people with agrammatism. *Clin Linguist Phon* 2009; **23**: 133-155 [PMID: [19197582](https://pubmed.ncbi.nlm.nih.gov/19197582/) DOI: [10.1080/02699200802602985](https://doi.org/10.1080/02699200802602985)]
- 114 **Tao J**, Kang Y, Li A. Prosody conversion from neutral speech to emotional speech. *IEEE Trans Audio Speech Lang Process* 2006; **14**: 1145-1154 [DOI: [10.1109/TASL.2006.876113](https://doi.org/10.1109/TASL.2006.876113)]
- 115 **Scherer KR**. Vocal affect expression: A review and a model for future research. *Psychol Bull* 1986; **99**: 143-165 [PMID: [3515381](https://pubmed.ncbi.nlm.nih.gov/3515381/) DOI: [10.1037//0033-2909.99.2.143](https://doi.org/10.1037//0033-2909.99.2.143)]
- 116 **Davitz JR**, Beldoch M. The Communication of Emotional Meaning. New York: McGraw-Hill, 1964
- 117 **Rabiner LR**, Schafer RW. Digital processing of speech signals. New Jersey: Prentice Hall, 1978: 121-123
- 118 **Hernando J**, Nadeu C. Linear prediction of the one-sided autocorrelation sequence for noisy speech recognition. *IEEE Trans Speech Audio Process* 1997; **5**: 80-84 [DOI: [10.1109/89.554273](https://doi.org/10.1109/89.554273)]
- 119 **Le Bouquin R**. Enhancement of noisy speech signals: Application to mobile radio communications. *Speech Commun* 1996; **18**: 3-19 [DOI: [10.1016/0167-6393\(95\)00021-6](https://doi.org/10.1016/0167-6393(95)00021-6)]
- 120 **Bou-Ghazale SE**, Hansen JH. A comparative study of traditional and newly proposed features for recognition of speech under stress. *IEEE Trans Speech Audio Process* 2000; **8**: 429-442 [DOI: [10.1109/89.848224](https://doi.org/10.1109/89.848224)]
- 121 **Rabiner LR**. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc IEEE* 1989; **77**: 257-286
- 122 **Yoon BJ**. Hidden Markov Models and their Applications in Biological Sequence Analysis. *Curr Genomics* 2009; **10**: 402-415 [PMID: [20190955](https://pubmed.ncbi.nlm.nih.gov/20190955/) DOI: [10.2174/138920209789177575](https://doi.org/10.2174/138920209789177575)]
- 123 **Duda RO**, Hart PE, Stork DG, Ionescu A. Pattern classification, chapter nonparametric techniques. Wiley-Interscience Publication, 2000
- 124 **Haykin S**. Neural networks and learning machines, 3/E. Pearson Education India, 2010



Published by **Baishideng Publishing Group Inc**
7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA

Telephone: +1-925-3991568

E-mail: bpgoffice@wjgnet.com

Help Desk: <https://www.f6publishing.com/helpdesk>

<https://www.wjgnet.com>

