









# An Entropy-Based Computational Classifier for Positive and Negative Emotions in Voice Signals

A. D. Herrera-Ortiz<sup>1</sup> , G. A. Yáñez-Casas<sup>2,4</sup> ,  
J. J. Hernández-Gómez<sup>2,4</sup> , M. G. Orozco-del-Castillo<sup>3,4</sup> ,  
M. F. Mata-Rivera<sup>1</sup> , and R. de la Rosa-Rábago<sup>2</sup> 

<sup>1</sup> Unidad Profesional Interdisciplinaria en Ingeniería y Tecnologías Avanzadas, Sección de Estudios de Posgrado e Investigación, Instituto Politécnico Nacional, Mexico City, Mexico

<sup>2</sup> Centro de Desarrollo Aeroespacial, Instituto Politécnico Nacional, Mexico City, Mexico

[jjhernandezgo@ipn.mx](mailto:jjhernandezgo@ipn.mx)

<sup>3</sup> Departamento de Sistemas y Computación, Tecnológico Nacional de México/IT de Mérida, Mérida, Yucatán, Mexico

<sup>4</sup> AAAI Student Chapter at Yucatán, México (AAAIMX), Association for the Advancement of Artificial Intelligence, Mexico City, Mexico

**Abstract.** The detection, classification and analysis of emotions has been an intense research area in the last years. Most of the techniques applied for emotion recognition are those comprised by Artificial Intelligence, such as neural networks, machine learning and deep learning, which are focused on the training and learning of models. In this work, we propose a rather different approach to the problem of detection and classification of emotion within voice speech, regarding sound files as information sources in the context of Shannon's information theory. By computing the entropy content of each audio, we find that emotion in speech can be classified into two subsets: positive and negative. To be able to perform the entropy computation, we first compute the Fourier transform to digital audio recordings, bearing in mind that the voice signal has a bandwidth 100 Hz and 4 kHz. The discrete Fourier spectrum is then used to set the alphabet and then the occurrence probabilities of each symbol (frequency) is used to compute the entropy for non-hysterical information sources. A dataset consisting of 1,440 voice audios performed by professional voice actors was analysed through this methodology, showing that in most cases, this simple approach is capable of performing the positive/negative emotion classification.

**Keywords:** Emotion analysis · Pattern recognition · Computational entropy · Information theory · Information source · Fourier transform · Frequency alphabet · Voice signals · Sound · Speech

## 1 Introduction

The identification, classification and analysis of emotions is a fertile, active and open research area within the pattern recognition field. Historically, the widest source of information to perform emotion detection has been text. However, a remarkable surge in the availability of text sources for sentiment analysis arrived in the last two decades with the massive spreading of Internet [1]. Moreover, the arisal of web-based social networks, particularly designed for the social interaction, eased their usage for the sharing of sentiments, generating massive amounts of information to be mined for the comprehension of human psyche [1,2]. Traditionally, as emotions detection and classification has been performed (mostly on text sources) with different techniques of Artificial Intelligence (AI), sentiment analysis is commonly regarded as an area of this same field [3].

Furthermore, the rise of social networks also allowed people to find new ways of expressing their emotions, with the use of content like emoticons, pictures as memes, audio and video [1,4], showing the necessity of generating methods to expand the sentiment analysis to this novel sources of information. Accordingly, much research has been performed in the field of emotion analysis within social networks content, which is mostly based on the analysis of text/comments with AI techniques [5–11]. Several applications in this field are healthcare [12,13], social behavioural assessment [14,15], touristic perception [16], identification of trends in conflicting versus non-conflicting regions [17], evaluation of influence propagation models on social networks [18], emotions identification in text/emoticons/emojis [19], among many others. A review on textual emotion analysis can be found in [20].

With respect to images, the analysis has been focused on facial emotion recognition mainly through the combination of AI and techniques of digital image processing [21]. In [22] a 2D canonical correlation was implemented, [23] combines the distance in facial landmarks along with a genetic algorithm, [24] used a deep learning approach to identify the motions of painters with their artwork, [25] used the maximum likelihood distributions to detect neutral, fear, pain, pleasure and laugh expressions in video stills, [26] uses a multimodal Graph Convolutional Network to perform a conjoint analysis of aesthetic and emotion feelings within images, [27] uses improved local binary pattern and wavelet transforms to assess the learning states and emotions of students in online learning, [28] uses principal component analysis and deep learning methods to identify emotion in children with autism spectrum disorder, [29] used facial thermal images, deep reinforcement learning and IoT robotic devices to assess attention-deficit hyperactivity disorder in children, while [30] fuzzifies emotion categories in images to assess them through a deep metric learning. A recent review on the techniques for emotion detection in facial images is reported in [31].

However, a much lesser studied area within emotion recognition is the emotion analysis within audio sources, specifically in voice/speech. The first attempts were based on the classification of emotions by parameters of the audio signal; for instance, for english and malayan voices and for six emotions, the average pitch, the pitch range and jitter were assessed for signals of both male and

female voices, finding that the language does not affect the emotional speech [32], while [33] applied data mining algorithms to prosody parameters extracted from non-professional voice actors. Also, [34] extracted 65 acoustic parameters to assess anger, contempt, fear, happiness, interest, lust, neutral, pride, relief, sadness, and shame emotional stages in over 100 professional actors from five English-speaking countries. Later, medical technology was applied using functional magnetic resonance images to measure the brain activity while the patient was giving a speech which in turn was recorded and computer-processed [35]. More algorithmical approaches were developed later, such as fuzzy logic reasoners [36, 37], discriminant analysis focused on nursing experience [38], the use of statistical similarity measurements to categorise sentiments in acted, natural and induced speeches [39, 40], the use of subjective psychological criteria to improve voice database design, parametrisation and classification schemes [41], among others. Machine learning approaches have also been developed, as the recognition of positive, neutral and negative emotions on spontaneous speech in children with autism spectrum disorders through support vector machines [42], the application of the k-nearest neighbour method to signal parameters as pitch, temporal and duration on theatrical plays for identification of happy, angry, fear, and neutral emotions [43], the simultaneous use of ant colony optimisation and k-nearest neighbour algorithms to improve the efficiency of speech emotion recognition, focusing only on the spectral roll-off, spectral centroid, spectral flux, log energy, and formats at few chosen frequency sub-bands [44], as well as the real time analysis of TV debates' speech through a deep learning approach in the parameter space [45]. In the field of neural networks, a neurolinguistic processing model based on neural networks to conjointly analyse voice through the acoustic parameters of tone, pitch, rate, intensity, meaning, etc., along with text analysis based on linguistic features was developed by [46], while [47] proposes the use of a multi-layer perceptron neural network to classify emotions by the Mel frequency Cepstral Coefficient, its scaled spectrogram frequency, chroma and tonnetz parameters. Moreover, some studies suggest that, when available, the conjoint analysis of voice and facial expressions could lead to a better performance on emotion classification than the two techniques used separately [48].

As can be observed, there exist two main approaches to the problem of emotion analysis in voice/speech records, which can be used together: the direct analysis of parameters derived from the sound signal, and the use of AI techniques at many levels to build recognition and classification schemes. The main drawback of the systems based on AI methods is that they are subject to a training process that might be prone to bias and that highly depends on the training dataset, which might be inappropriately split [49]; moreover, the presence of hidden variables as well as mistaking the real objective are common drawbacks in the field [49, 50]. Collateral drawbacks are likely the large amount of time and computer resources required to train the AI-based systems. In [51] and [52] the subject of how to build representative AI models in general is explored.

In this work, we deviate from the traditional approaches to the problem of emotion analysis in order to explore a novel approach that regards the voice/speech recording as an information source in the framework of Shannon's

information theory [53]. In particular, we compute the information entropy of a voice/speech signal in order to classify emotion into two categories, positive and negative emotions, by generating an alphabet consisting on the frequency content of a human-audible sub-band. Although Shannon entropy has been previously used to perform pattern recognition in sound, it has been applied mainly to the heart sounds classification [54, 55]. The outcome shows that this approach is suitable for a very fast automatic classification of positive and negative emotions, which lacks of a training phase by its own nature. This work is organised as follows: in Sect. 2 we show the theoretical required background as well as the dataset under use, while in Sect. 3 we show the followed procedure along with the obtained results. Finally, in Sects. 4 and 5 we pose some final remarks as well as future possible paths to extend the presented work.

## 2 Materials and Methods

### 2.1 Frequency Domain Analysis

Since the inception of the analysis in the frequency domain by Joseph Fourier in 1882 [56], Fourier series for periodic waveforms and Fourier transform for non-periodic ones have been cornerstones of modern mathematical and numerical analysis. Fourier transforms place time series in the frequency domain, so they are able to provide their frequency content. Moreover, both continuous and discrete waveforms are likely to be analysed through Fourier analysis. In this work, we focus on discrete time series because most of audio sources available nowadays are binary files stored, processed and transmitted in digital computers. Let  $x(n)$  be a discrete time series of finite energy, its Fourier transform is given by

$$X(w) = \sum_{n=-\infty}^{\infty} x(n)e^{-jwn}, \quad (1)$$

where  $X(w)$  represents the frequency spectrum of  $x(n)$  [57]. Such frequency content allows to classify the signal according to its power/energy density spectra, which are quantitatively expressed as the bandwidth. Fourier transform has been successfully applied for more than a century, in virtually any field of knowledge as it can be imagined for signal analysis, such as in medicine [58–61], spectroscopy/spectrometry [60–69], laser-material interaction [70], image processing [59, 71–73], big data [74], micro-electro-mechanical systems (MEMS) devices [75], food technology [73, 76, 77], aerosol characterisation and assessment [78], vibrations analysis [79], chromatic dispersion in optical fiber communications [80], analysis of biological systems [81], characterisation in geological sciences [82], data compression [83], catalyst surface analysis [84], profilometry [85], among several others.

Frequency domain analysis can be applied to any signal from which information is to be extracted. In the case of the voice signal herein studied, the bandwidth is limited to a frequency range 100 Hz and 4 kHz.

## 2.2 Shannon's Entropy Computation

The fundamental problem of communications, i.e. to carry entirely a message from one point to another, was first posed mathematically in [53]. Within his theory, messages are considered discrete in the sense that they might be represented by a number of symbols, regardless of the continuous or discrete nature of the information source, because any continuous source should be eventually discretised in order to be further transmitted. The selected set of symbols to represent certain message is called the alphabet, so that an infinite number of messages could be coded by such alphabet, regardless of its finitude.

In this sense, different messages coded in the same alphabet use different symbols, so the probability of appearance of each could vary from each one to the other. Therefore, the discrete source of information could be considered as a stochastic process, and conversely, any stochastic process that produces a discrete sequence of symbols selected from a finite set will be a discrete source [53]. An example of this is the digital voice signal.

For a discrete source of information in which the probabilities of occurrence of events are known, there is a measure of how much choice is involved in selecting the event or how uncertain we are about its outcome. According to theorem 2 of [53], there should be a function  $H$  that satisfies the properties of being continuous on the probabilities of the events ( $p_i$ ), of being a monotonically increasing function of  $n$  and as well as being additive. The logarithmic function meets such requirements, and it is optimal for considering the influence of the statistics of very large messages in particular, as the occurrence of the symbols tends to be very large. In particular, base 2 logarithms are singularly adequate to measure the information, choice and uncertainty content of digital (binary) coded messages. Such a function then takes the form

$$H = -K \sum_{i=1}^n p_i \log(p_i), \quad (2)$$

where the positive constant  $K$  sets a measurement unit and  $n$  is the number of symbols in the selected alphabet.  $H$  is the so called information (or Shannon's) entropy for a set of probabilities  $p_1, \dots, p_n$ . It must be noted that the validity of Eq. 2 relies in the fact that each symbol within the alphabet is equiprobable, i.e. for information sources that do not possess hysteresis processes. For alphabets with symbols that are not equally probable, Eq. 2 is modified, yielding the conditional entropy.

Beyond the direct applications of information entropy in the field of communications systems, it has been also used for assessment and classification purposes. For instance, [86] evaluates non-uniform distribution of assembly features in precision instruments, [87] applies it to multi-attribute utility analysis, [88] uses a generalised maximum entropy principle to identify graphical ARMA models, [89] studies critical characteristics of self-organised behaviour of concrete under uniaxial compression, [90] explores interactive attribute reduction for unlabelled mixed data, [91] improves neighbourhood entropies for uncertainty analysis and

intelligent processing, [92] proposes an inaccuracy fuzzy entropy measure for a pair of probability distribution and discuss its relationship with mean codeword length, [93] develops proofs for quantum error-correction codes, [94] performs attribute reduction for unlabelled data through an entropy based missclassification cost function, [95] applies the cross entropy of mass function in similarity measure, [96] detects stress-induced sleep alteration in electroencephalographic records, [97] uses entropy and symmetry arguments to assess the self-replication problem in robotics and artificial life, [98] applies it to the quantisation of local observations for distributed detection, etc.

Although the information entropy has been largely applied to typical text sources, the question of how to apply it to digital sound sources could entail certain difficulties, as the definition of an adequate alphabet in order to perform the information measurement. In this work, we first implement a fast Fourier transform algorithm to a frequency band of digital voice audios in order to set a frequency alphabet. The symbols of such alphabet are finite as the sound files are sampled to the same bitrate.

### 2.3 Voice Signals Dataset

The dataset used in this work was obtained from the Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) [99], and it consists of 1,440 audio-only WAV files, performed of 24 professional voice actors (12 women, 12 men), who vocalise English matching statements with a neutral American accent.

Each actor performs calm, happy, sad, angry, fearful, surprised, disgusted, and neutral expressions. Each expression is performed at two levels of emotional intensities: normal and loud. Each actor vocalised two different statements: 1 stands for “Kids are talking by the door” and 2 for “Dogs are sitting by the door”. Finally, each vocalisation was recorded twice which is stated as 1<sup>st</sup> or 2<sup>nd</sup> repetition. Each audio is approximately 3 s long. Table 1 shows the classification of the voice dataset, in which it can be observed that such emotions have been separated into two subsets of positive and negative emotions.

**Table 1.** Number of files within the voice dataset, distributed by emotions, which are separated in positive and negative emotions subsets.

		Emotions														Total				
		Positive						Negative						Neutral						
		<i>Happy</i>		<i>Surprised</i>		<i>Calm</i>		<i>Sad</i>		<i>Fearful</i>		<i>Disgusted</i>		<i>Angry</i>						
Message	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2				
Loud	1 <sup>st</sup>	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	-	-	336
	2 <sup>nd</sup>	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	-	-	336
Normal	1 <sup>st</sup>	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	384
	2 <sup>nd</sup>	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	384
Total		96	96	96	96	96	96	96	96	96	96	96	96	96	96	96	96	48	48	1440
		192		192		192		192		192		192		192		96				

### 3 Development and Results

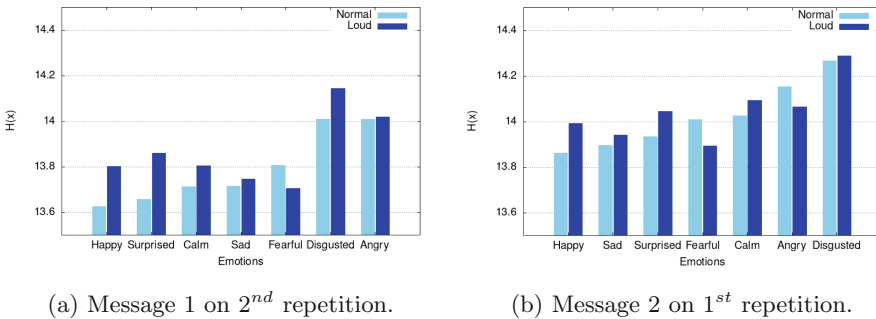
The methods in the aforementioned section were implemented with the Python programming language. First, the librosa library, which is focused on the processing of audio and music signals [100], was implemented to obtain the sampled data of the audio at a sampling bitrate of 22,050 Hz. In order to obtain the frequency spectrum, the Fast Fourier Transform (FFT) algorithm was used to perform the discrete fourier transform. FFT was implemented through the SciPy library, a collection of mathematical algorithms and functions built on top of the NumPy extension to Python, adding significant enhancement by providing high-level commands and classes for manipulating and displaying data [101]. SciPy is a system prototyping and data processing environment that rivals systems like MATLAB, IDL, Octave, R-Lab, and SciLab [101].

As the audios were all sampled to a bitrate of  $br = 22,050$  Hz, if their duration is of  $t$  s, then the number of time samples they possess is just  $br \cdot t$ , which conform a  $br \cdot t$ -sized vector. Then, the `scipy.fft.rfft` function is used to compute the 1D discrete Fourier transform (DFT) of  $n$  points of such real-valued vector [102].

In order to compute the value of entropy of each voice source with the described alphabet, the probability of each symbol (frequency values available in the Fourier spectra) is computed and then the entropy through Eq. (2) is finally calculated, as each frequency does not depend on the occurrence of another, i.e. the sound information source can be regarded as non-hysterical.

#### 3.1 Entropy Analysis

After the processing of the data set (Table 1), the entropy outcome of each audio was analysed for the following emotions: calm, happy, sad, angry, afraid, surprised and disgusted. As neutral expression does not express any emotion intentionally, it was discarded from the analysis. Due to the length of the dataset, here we only show some representative graphics of the obtained results. We also display the average results in what follows.



**Fig. 1.** Comparison of the average entropy for all actors on each emotion.

The first analysis is over the average of the entropy of all the actors comparing the loud against the normal emotional intensities. Results can be observed for both messages in Fig. 1, where the results are presented in ascending order with respect to normal emotional intensity values of entropy.

Table 2 shows in detail the average values of the entropy obtained from all of the 24 actors, including the cases of Fig. 1, classified according to intensity (loud and normal). The values are in ascending order in accordance with normal intensity. The values obtained in each repetition (1st and 2nd) of the messages are shown separately.

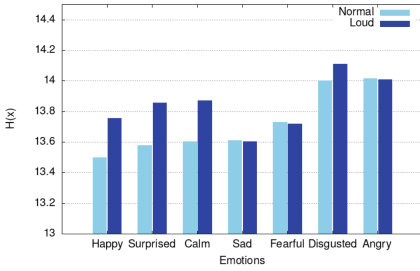
**Table 2.** Average entropy of all actors comparing intensity.

Message 1			Message 2		
Emotion	Normal	Loud	Emotion	Normal	Loud
1 <sup>st</sup> Repetition					
<i>Happy</i>	13.62937	13.78712	<i>Happy</i>	13.86193	13.99179
<i>Surprise</i>	13.67065	13.93201	<i>Sad</i>	13.89544	13.94124
<i>Sad</i>	13.72977	13.65707	<i>Surprise</i>	13.93365	14.04495
<i>Calm</i>	13.74445	13.91538	<i>Fear</i>	14.00737	13.89378
<i>Fear</i>	13.78522	13.72170	<i>Calm</i>	14.02606	14.09250
<i>Angry</i>	13.99452	14.01461	<i>Angry</i>	14.15280	14.06335
<i>Disgusted</i>	14.04117	14.18401	<i>Disgusted</i>	14.26644	14.28761
2 <sup>nd</sup> Repetition					
<i>Happy</i>	13.62500	13.80204	<i>Surprise</i>	13.87808	14.07231
<i>Surprise</i>	13.65754	13.85815	<i>Happy</i>	13.91507	13.94989
<i>Calm</i>	13.71306	13.80327	<i>Sad</i>	13.96764	13.93037
<i>Sad</i>	13.71528	13.74593	<i>Fear</i>	13.99382	13.84644
<i>Fear</i>	13.80536	13.70516	<i>Calm</i>	14.03019	14.12940
<i>Disgusted</i>	14.00738	14.14444	<i>Angry</i>	14.19368	14.12245
<i>Angry</i>	14.00800	14.01744	<i>Disgusted</i>	14.22615	14.33484

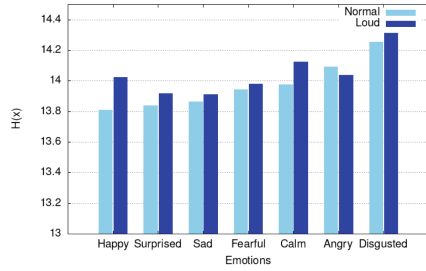
In order to explore the entropic content of the audios by gender, we compared the loud and normal intensities for both gender, whose graphics are in ascending order with respect to normal intensity, here shown in Figs. 2a and 2b for men and Figs. 2c and 2d for women.

Table 3a shows in detail the average values of the entropy obtained from the 12 male interpreters, as well as in Table 3b for women, which include the cases shown in Fig. 2. Both tables classify the entropy values according to intensity (normal and loud) in ascending order with respect to normal intensity, featuring separately the values of the 1st and 2nd repetitions of the messages.

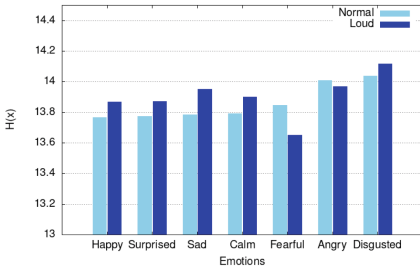




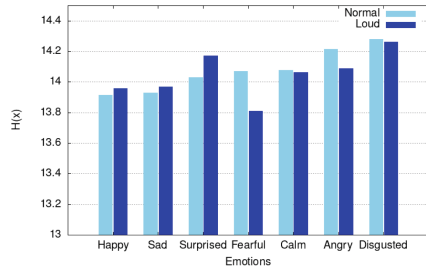
(a) Men’s message 1 on 1<sup>st</sup> repetition.



(b) Men’s message 2 on 1<sup>st</sup> repetition.



(c) Women’s message 1 on 2<sup>nd</sup> repetition.



(d) Women’s message 2 on 1<sup>st</sup> repetition.

**Fig. 2.** Normal vs loud intensity for both gender interpreters.

**Table 3.** Average entropy for both genders and messages, comparing intensity.

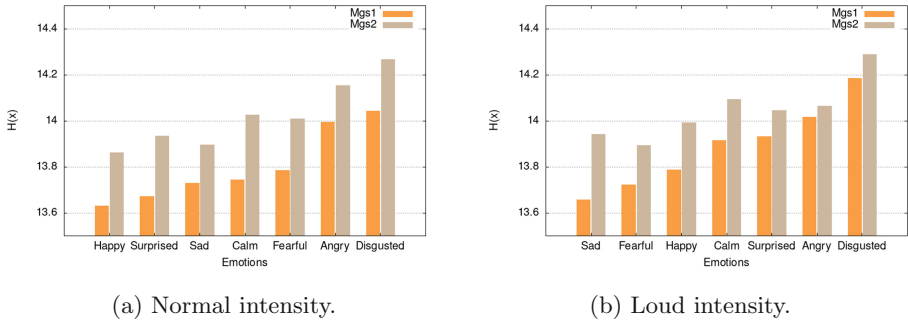
(a) Male interpreters					
Message 1			Message 2		
Emotion	Normal	Loud	Emotion	Normal	Loud
1 <sup>st</sup> Repetition					
<i>Happy</i>	13.49931	13.75591	<i>Happy</i>	13.80846	14.02408
<i>Surprised</i>	13.57929	13.85501	<i>Surprised</i>	13.83677	13.91841
<i>Calm</i>	13.60234	13.87002	<i>Sad</i>	13.86316	13.91226
<i>Sad</i>	13.60938	13.60310	<i>Fearful</i>	13.94512	13.97890
<i>Fearful</i>	13.73046	13.72030	<i>Calm</i>	13.97587	14.12232
<i>Disgusted</i>	14.00011	14.11143	<i>Angry</i>	14.09121	14.03747
<i>Angry</i>	14.01447	14.00702	<i>Disgusted</i>	14.25447	14.31288
2 <sup>nd</sup> Repetition					
<i>Happy</i>	13.48353	13.73792	<i>Surprised</i>	13.79324	13.97323
<i>Surprised</i>	13.54198	13.84495	<i>Happy</i>	13.86438	13.93191
<i>Calm</i>	13.63455	13.70720	<i>Fearful</i>	13.90677	13.91836
<i>Sad</i>	13.64754	13.54204	<i>Sad</i>	13.94234	13.83481
<i>Fearful</i>	13.76416	13.75924	<i>Calm</i>	13.96887	14.02359
<i>Disgusted</i>	13.97689	14.17172	<i>Disgusted</i>	14.14322	14.39701
<i>Angry</i>	14.00588	14.06509	<i>Angry</i>	14.15684	14.10251

(continued)

**Table 3.** (continued)

(b) Female interpreters					
Message 1			Message 2		
Emotion	Normal	Loud	Emotion	Normal	Loud
<i>1<sup>st</sup> Repetition</i>					
<i>Happy</i>	13.75942	13.81833	<i>Happy</i>	13.91540	13.95950
<i>Surprised</i>	13.76201	14.00901	<i>Sad</i>	13.92772	13.97022
<i>Fearful</i>	13.83999	13.72310	<i>Surprised</i>	14.03053	14.17150
<i>Sad</i>	13.85016	13.71104	<i>Fearful</i>	14.06962	13.80866
<i>Calm</i>	13.88657	13.96074	<i>Calm</i>	14.07626	14.06269
<i>Angry</i>	13.97456	14.02220	<i>Angry</i>	14.21440	14.08923
<i>Disgusted</i>	14.08223	14.25659	<i>Disgusted</i>	14.27841	14.26233
<i>2<sup>nd</sup> Repetition</i>					
<i>Happy</i>	13.76648	13.86615	<i>Surprised</i>	13.96292	14.17139
<i>Surprised</i>	13.77310	13.87135	<i>Happy</i>	13.96576	13.96787
<i>Sad</i>	13.78301	13.94982	<i>Sad</i>	13.99293	14.02593
<i>Calm</i>	13.79157	13.89934	<i>Fearful</i>	14.08087	13.77452
<i>Fearful</i>	13.84656	13.65107	<i>Calm</i>	14.09151	14.23521
<i>Angry</i>	14.01012	13.96978	<i>Angry</i>	14.23052	14.14239
<i>Disgusted</i>	14.03787	14.11715	<i>Disgusted</i>	14.30908	14.27268

In what follows, we present the results of the average entropy for message 1 against message 2, for the same motional intensity. The general results for the 24 actors are observed in Fig. 3, where each plot is ordered in ascending order with respect to message 1.



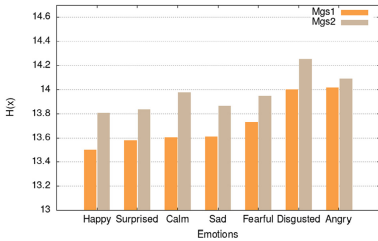
**Fig. 3.** Message 1 vs 2 on *1<sup>st</sup>* repetition.

Likewise, Table 4 shows all the average values of the entropy obtained from the 24 actors, including the cases shown in Fig. 3, classified in accordance to the type of message (1 or 2). The values are presented in ascending order with respect to the normal intensity. The values obtained in each repetition (1st and 2nd) of the messages are clearly separated.

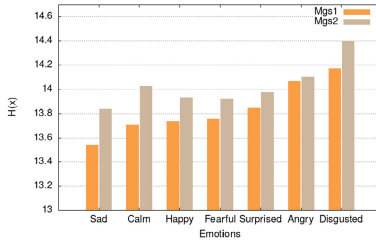
**Table 4.** Average entropy of all actors comparing type of message.

Emotion	Normal		Emotion	Loud	
	Mgs 2	Mgs 1		Mgs 2	Mgs 1
<b>1<sup>st</sup> Repetition</b>					
<i>Happy</i>	13.86193	13.62937	<i>Sad</i>	13.94124	13.65707
<i>Surprise</i>	13.93365	13.67065	<i>Fear</i>	13.89378	13.72170
<i>Sad</i>	13.89544	13.72977	<i>Happy</i>	13.99179	13.78712
<i>Calm</i>	14.02606	13.74445	<i>Calm</i>	14.09250	13.91538
<i>Fear</i>	14.00737	13.78522	<i>Surprise</i>	14.04495	13.93201
<i>Angry</i>	14.15280	13.99452	<i>Angry</i>	14.06335	14.01461
<i>Disgusted</i>	14.26644	14.04117	<i>Disgusted</i>	14.28761	14.18401
<b>2<sup>nd</sup> Repetition</b>					
<i>Happy</i>	13.91507	13.62500	<i>Fear</i>	13.84644	13.70516
<i>Surprise</i>	13.87808	13.65754	<i>Sad</i>	13.93037	13.74593
<i>Calm</i>	14.03019	13.71306	<i>Happy</i>	13.94989	13.80204
<i>Sad</i>	13.96764	13.71528	<i>Calm</i>	14.12940	13.80327
<i>Fear</i>	13.99382	13.80536	<i>Surprise</i>	14.07231	13.85815
<i>Disgusted</i>	14.22615	14.00738	<i>Angry</i>	14.12245	14.01744
<i>Angry</i>	14.19368	14.00800	<i>Disgusted</i>	14.33484	14.14444

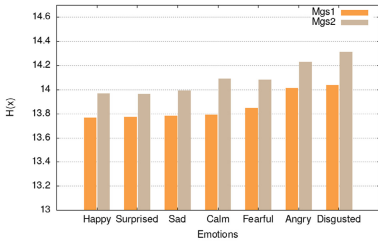
Moreover, the average entropy comparison between messages 1 and 2 (with the same emotional intensity) is then shown for men in Figs. 4a and 4b and in Figs. 4c and 4d for women. Each graph is ordered in ascending order with respect to message 1.



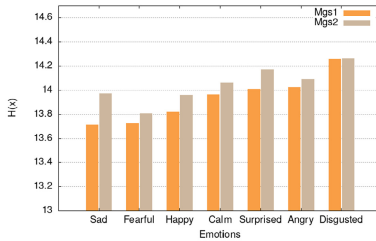
(a) Men’s normal intensity on 1<sup>st</sup> repetition.



(b) Men’s loud intensity on 2<sup>nd</sup> repetition.



(c) Women’s normal intensity on 2<sup>nd</sup> repetition.



(d) Women’s loud intensity on 1<sup>st</sup> repetition.

**Fig. 4.** Message 1 vs 2 with the same intensity for both gender interpreters.

Table 5a shows in detail the average values of the entropy obtained from the 12 male interpreters, classified according to the message 1 or 2. The values are in ascending order with respect to the normal intensity. The values obtained in each repetition (1st and 2nd) of the messages were separated. The exact same setup for the average values of the entropy obtained from the 12 female interpreters can be observed in Table 5b. The cases shown in Fig. 4 are also include here.

**Table 5.** Average entropy for both gender actors comparing type of message.

	Normal			Loud	
Emotion	Mgs 2	Mgs 1	Emotion	Mgs 2	Mgs 1
(a) Male interpreters					
1 <sup>st</sup> Repetition					
<i>Happy</i>	13.97587	13.49931	<i>Sad</i>	14.02408	13.60310
<i>Surprise</i>	13.86316	13.57929	<i>Fear</i>	14.03747	13.72030
<i>Calm</i>	13.80846	13.60234	<i>Happy</i>	14.12232	13.75591
<i>Sad</i>	13.94512	13.60938	<i>Surprise</i>	13.97890	13.85501
<i>Fear</i>	14.25447	13.73046	<i>Calm</i>	14.31288	13.87002
<i>Disgusted</i>	13.81052	14.00011	<i>Angry</i>	13.91841	14.00702
<i>Angry</i>	13.83677	14.01447	<i>Disgusted</i>	13.91226	14.11143
2 <sup>nd</sup> Repetition					
<i>Happy</i>	13.94234	13.48353	<i>Sad</i>	13.83481	13.54204
<i>Surprise</i>	13.96887	13.54198	<i>Calm</i>	14.02359	13.70720
<i>Calm</i>	14.15684	13.63455	<i>Happy</i>	13.93191	13.73792
<i>Sad</i>	13.86438	13.64754	<i>Fear</i>	13.91836	13.75924
<i>Fear</i>	13.79324	13.76416	<i>Surprise</i>	13.97323	13.84495
<i>Disgusted</i>	13.90677	13.97689	<i>Angry</i>	14.10251	14.06509
<i>Angry</i>	14.14322	14.00588	<i>Disgusted</i>	14.39701	14.17172
(b) Female interpreters					
1 <sup>st</sup> Repetition					
<i>Happy</i>	14.07626	13.75942	<i>Sad</i>	14.17150	13.71104
<i>Surprise</i>	13.91540	13.76201	<i>Fear</i>	13.80866	13.72310
<i>Fear</i>	14.03053	13.83999	<i>Happy</i>	14.26233	13.81833
<i>Sad</i>	14.21440	13.85016	<i>Calm</i>	14.08923	13.96074
<i>Calm</i>	13.88075	13.88657	<i>Surprise</i>	13.97022	14.00901
<i>Angry</i>	14.06962	13.97456	<i>Angry</i>	14.06269	14.02220
<i>Disgusted</i>	13.92772	14.08223	<i>Disgusted</i>	13.95950	14.25659
2 <sup>nd</sup> Repetition					
<i>Happy</i>	14.08087	13.76648	<i>Fear</i>	14.23521	13.65107
<i>Surprise</i>	13.99293	13.77310	<i>Happy</i>	14.14239	13.86615
<i>Sad</i>	14.09151	13.78301	<i>Surprise</i>	13.96787	13.87135
<i>Calm</i>	13.96576	13.79157	<i>Calm</i>	14.27268	13.89934
<i>Fear</i>	14.23052	13.84656	<i>Sad</i>	14.02593	13.94982
<i>Angry</i>	13.88605	14.01012	<i>Angry</i>	13.77452	13.96978
<i>Disgusted</i>	14.30908	14.03787	<i>Disgusted</i>	14.17139	14.11715

## 4 Discussion and Conclusions

In this work, a different approach to the analysis of emotion was proposed, since instead of applying the widely common methods of AI, a classification of emotions into positive and negative categories within speech sources is proposed through a tool of the theory of the information: the frequency-based Shannon's entropy. In order to compute information entropy, an alphabet based on the frequency symbols generated by the decomposition of original audio time series through the FFT algorithm is generated. Then, the probability of appearance of each frequency symbol is obtained from the Fourier spectrum so to finally compute the non-hysterical Shannon's entropy (see Eq. (2)).

As already mentioned in Sect. 1, the typical sources of information in which entropy calculation is performed are texts where the average entropy value ranges between 3 and 4. However, as it was observed in the average values herein provided, they range between 13 and 15. This is clearly due to the nature of the alphabet developed here. Given that in the texts, alphabets are composed of a number of symbols of the order of tens, they yield small values of entropy. However, in the frequency domain, sound signals generate much larger alphabets, yielding average entropies for a voice signal that are considerably higher than that of a text. It is also clear that if richer sound sources would be analysed through this method, as music files and not only speech, they would certainly yield larger values of entropy. It must be considered that a value of about 14 is much greater than the typical values of text entropy of 3–4, given the logarithmic function that characterises the computation of entropy.

As it can be observed through Sect. 3, a general tendency of positive emotions (happy, surprise and calm) to have lower values of entropy than the negative emotions (sad, fear, disgust and anger) is present. Thus, large values of entropy generally characterise negative emotions while lower values are typical of positive emotions, allowing to perform a pre-classification of emotions into these two categories, without the necessity of going to a training phase as in general machine learning algorithms.

**Table 6.** Average entropy classified by positive and negative emotion according to intensity.

Case Study	Positive Emotion		Negative Emotion	
	Normal	Loud	Normal	Loud
<i>All actors</i>	13.80708	13.94823	13.98687	13.97565
<i>Men</i>	13.71571	13.89704	13.94074	13.96713
<i>Women</i>	13.89846	13.99942	14.03300	13.98416

In order to better grasp the main result, in Table 6 we feature the average entropy values according to the (normal and loud) intensities, by considering both the positive and negative categories of emotions covered by this work.

Such values are presented for all of the actors, as well as separated by gender. It can be clearly observed that for the normal intensity, for the three averages (for all actors, for men and for women), positive emotions yield smaller values of entropy than those given by the negative emotions.

It is important to remember that for an information source within the context of Shannon's theory of information, symbols with lower probabilities to occur, are the ones that represent more information, since the number of symbols required to represent a certain message is less. If the symbols are more equiprobable for a message, the entropic content will be small. In other words, when there are symbols that occur infrequently, the entropic content would be higher. Also, the longitude of the message plays an important role since for short messages, the entropy will vary compared to a long ones, where the entropy will tend to stabilise [103]. In this sense, it can be clearly observed from Table 6 that the entropy values for the same intensity, comparing men against women, turn out larger for women, in both positive and negative emotions (normal intensity). This fact is consistent with the previous observation, because women in general excite a narrower frequency bandwidth, thus making more unfrequently symbols available, yielding to larger values of entropy.

The same pattern from the normal intensity is observed for the loud intensity for the average of all actors as well as for the average of men (see Table 6). It should be noted that for the loud intensities, the gap between the positive and negative emotions is smaller than for the normal intensity of the message. This is clearly because at a loud intensity, the amplitude of the time series is larger, thus in general increasing the probabilities of occurrence of each symbol. The particular case of female interpreters in which the loud intensity has a lower value for the negative emotion than for the positive emotion is likely because when women shout, they narrow their voices' frequency bandwidth, thus yielding less symbols with larger probabilities. This is not the case of male interpreters, that when they shout, tend to excite a larger portion of the spectrum, yielding lower values of entropy.

On the other side, Table 7 also shows the average values of entropy for the positive and negative categories of emotions, but according to the type of message. It could be noted that the general results are coherent because in general, message 2 has larger values of entropy than message 1. This could be explained subjectively because people could tend to be more expressive with his emotions when talking about animals (dogs, in the case of message 2) than when talking about kids (message 1). Moreover, the high values of entropy for message 2 could be due to the fact that naturally, persons are more susceptible to get negative emotions to animals. These facts could have influenced the actors when vocalising message 2 with respect to message 1. Moreover, Table 7 confirms that for all the cases (all of the actors, the men and the women), the entropy values of positive emotions are lower than the values of entropy for negative emotions, regardless of the analysed message, confirming the results shown in Table 6.

**Table 7.** Average entropy classified by positive and negative emotion according to message.

Case Study	Positive Emotion		Negative Emotion	
	Message 1	Message 2	Message 1	Message 2
<i>All actors</i>	13.76150	13.99381	13.89231	14.07021
<i>Men</i>	13.67599	13.93675	13.85805	14.04983
<i>Women</i>	13.84700	14.05087	13.92657	14.09059

Various sound classification applications using AI techniques are based on the implementation of neural network variants, such as Deep Neural Network (DNN) [104], Convolutional Neural Networks (CNN) [105, 106], Recurrent Neural Networks (RNN) [107], among others. Although the use of AI techniques allows predicting the behaviour of the data from an exhaustive training of the chosen algorithm, fixed parameters such as entropy always allows an analysis without estimates or predictions. Thus, entropy values gives a clear idea of the behaviour of the signal from itself, yielding a more reliable and direct result [108].

Although not directly related to information classification, entropy calculation is useful in the context of communication systems, as it represents a measure of the amount of information that can be transmitted. Parameters such as channel capacity, joint entropy, data transmission rate, error symbol count, among others, use entropy to be determined [53]. These parameters become important when the information already classified or processed needs to be transmitted. Various applications such as those exposed in [109] and [110] combine the classification of information with its use in communication systems, especially those that require direct interaction with humans. Despite Shannon's entropy has been previously used to perform pattern recognition in sound, it has been mainly applied to the heart sounds classification [54, 55], and not in the context herein studied.

As final remarks, in this work we find Shannon's information to be a reliable tool that is able to perform a very quick classification of emotions into positive and negative categories. The computation of entropy based on the Fourier frequency spectrum also allows to categorise a message considering the amplitude of the original time series (if it is vocalised in normal or loud manner) as well as into male and female broadcaster. However, as previously mentioned through this section, further experiments with larger speech datasets should be performed in order to find stabilised entropy values to pose limiting quantitative criteria. In this way, for its simplicity and quickness, this novel approach could also serve as a pre-classification system for emotions in order to prepare training datasets for more complex machine learning algorithms to perform finer classifications of sentiments.

## 5 Future Work

This research can be extended in the following pathways:

- To expand this analysis to longer voice records as complete speeches.
- To expand this proposal to perform emotion analyses in analogical voice signals.
- To extend this analysis to assess the entropic content of voice audios in languages different than English.
- To explore the entropic content of other sound sources as music.
- To complement this approach with further tools of information theory [53] and signal analysis techniques, in order to be able to perform a finer emotion classification.

**Acknowledgements.** This work was supported by projects SIP 20220907, 20222032, 20220378 and EDI grant, by Instituto Politécnico Nacional/Secretaría de Investigación y Posgrado, as well as by projects 13933.22-P and 14601.22-P from Tecnológico Nacional de México/IT de Mérida.

## References

1. Nandwani, P., Verma, R.: A review on sentiment analysis and emotion detection from text. *Social Network Analysis and Mining* **11**(1), 1–19 (2021). <https://doi.org/10.1007/s13278-021-00776-6>
2. Li, H.H., Cheng, M.S., Hsu, P.Y., Ko, Y., Luo, Z.: Exploring Chinese dynamic sentiment/emotion analysis with text Mining–Taiwanese popular movie reviews comment as a case. In: B. R., P., Thenkanidiyoor, V., Prasath, R., Vanga, O. (eds) *Mining Intelligence and Knowledge Exploration, MIKE 2019*. LNCS, vol. 11987. Springer, Cham (2020). [https://doi.org/10.1007/978-3-030-66187-8\\_9](https://doi.org/10.1007/978-3-030-66187-8_9)
3. Tyagi, E., Sharma, A.: An intelligent framework for sentiment analysis of text and emotions - a review. In: *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, pp. 3297–3302 (2018)
4. Fulse, S., Sugandhi, R., Mahajan, A.: A survey on multimodal sentiment analysis. *Int. J. Eng. Res. Technol.* **3**(11), 1233–1238 (2014)
5. Wu, D., Zhang, J., Zhao, Q.: A text emotion analysis method using the dual-channel convolution neural network in social networks. *Math. Probl. Eng.* **2020**, 1–10 (2020)
6. Lu, X., Zhang, H.: An emotion analysis method using multi-channel convolution neural network in social networks. *Comput. Model. Eng. Sci.* **125**(1), 281–297 (2020)
7. Chawla, S., Mehrotra, M.: A comprehensive science mapping analysis of textual emotion mining in online social networks. *Int. J. Adv. Comput. Sci. Appl.* **11**(5), 218–229 (2020)
8. Wickramaarachchi, W., Kariapper, R.: An approach to get overall emotion from comment text towards a certain image uploaded to social network using latent semantic analysis. In: *2017 2nd International Conference on Image, Vision and Computing (ICIVC)*, pp. 788–792 (2017)



9. Jamaluddin, M., Abidin, S., Omar, N.: Classification and quantification of user's emotion on Malay language in social network sites using latent semantic analysis. In: 2016 IEEE Conference on Open Systems (ICOS), pp. 65–70 (2017)
10. Iglesias, C., Sánchez-Rada, J., Vulcu, G., Buitelaar, P.: Linked data models for sentiment and emotion analysis in social networks. Elsevier (2017)
11. Colnerič, N., Demšar, J.: Emotion recognition on twitter: comparative study and training a unison model. *IEEE Trans. Affect. Comput.* **11**(3), 433–446 (2018)
12. Li, T.S., Gau, S.F., Chou, T.L.: Exploring social emotion processing in autism: evaluating the reading the mind in the eyes test using network analysis. *BMC Psychiatry* **22**(1), 161 (2022)
13. Jiang, S.Y., et al.: Network analysis of executive function, emotion, and social anhedonia. *PsyCh J.* **11**(2), 232–234 (2022)
14. Yu, J.: Research on key technologies of analysis of user emotion fluctuation characteristics in wireless network based on social information processing. In: Liu, S., Ma, X. (eds.) ADHIP 2021. LNICST, vol. 416, pp. 142–154. Springer, Cham (2022). [https://doi.org/10.1007/978-3-030-94551-0\\_12](https://doi.org/10.1007/978-3-030-94551-0_12)
15. Han, Z.M., Huang, C.Q., Yu, J.H., Tsai, C.C.: Identifying patterns of epistemic emotions with respect to interactions in massive online open courses using deep learning and social network analysis. *Comput. Hum. Behav.* **122**, 106843 (2021)
16. Chen, X., Li, J., Han, W., Liu, S.: Urban tourism destination image perception based on LDA integrating social network and emotion analysis: the example of Wuhan. *Sustainability (Switzerland)* **14**(1), 12 (2022)
17. Wani, M., Agarwal, N., Jabin, S., Hussain, S.: User emotion analysis in conflicting versus non-conflicting regions using online social networks. *Telematics Inform.* **35**(8), 2326–2336 (2018)
18. Liu, X., Sun, G., Liu, H., Jian, J.: Social network influence propagation model based on emotion analysis. In: 2018 14th International Conference on Semantics, Knowledge and Grids (SKG), pp. 108–114 (2018)
19. Egorova, E., Tsarev, D., Surikov, A.: Emotion analysis based on incremental online learning in social networks. In: 2021 IEEE 15th International Conference on Application of Information and Communication Technologies (AICT) (2021)
20. Peng, S., et al.: A survey on deep learning for textual emotion analysis in social networks. *Digit. Commun. Netw.* (2021)
21. Gonzalez, R., Woods, R.: *Digital Image Processing*. Pearson (2018)
22. Ullah, Z., Qi, L., Binu, D., Rajakumar, B., Mohammed Ismail, B.: 2-D canonical correlation analysis based image super-resolution scheme for facial emotion recognition. *Multimedia Tools Appl.* **81**(10), 13911–13934 (2022)
23. Bae, J., Kim, M., Lim, J.: Emotion detection and analysis from facial image using distance between coordinates feature. In: 2021 International Conference on Information and Communication Technology Convergence (ICTC), vol. 2021, pp. 494–497 (2021)
24. Zhang, J., Duan, Y., Gu, X.: Research on emotion analysis of Chinese literati painting images based on deep learning. *Front. Psychol.* **12**, 723325 (2021)
25. Prossinger, H., Hladky, T., Binter, J., Boschetti, S., Riha, D.: Visual analysis of emotions using AI image-processing software: possible male/female differences between the emotion pairs “neutral”-“fear” and “pleasure”-“pain”. In: The 14th Pervasive Technologies Related to Assistive Environments Conference, pp. 342–346 (2021)
26. Miao, H., Zhang, Y., Wang, D., Feng, S.: Multi-output learning based on multi-modal GCN and co-attention for image aesthetics and emotion analysis. *Mathematics* **9**(12), 1437 (2021)

27. Wang, S.: Online learning behavior analysis based on image emotion recognition. *Traitement du Sign.* **38**(3), 865–873 (2021)
28. Sushma, S., Bobby, T., Malathi, S.: Emotion analysis using signal and image processing approach by implementing deep neural network. *Biomed. Sci. Instrum.* **57**(2), 313–321 (2021)
29. Lai, Y., Chang, Y., Tsai, C., Lin, C., Chen, M.: Data fusion analysis for attention-deficit hyperactivity disorder emotion recognition with thermal image and internet of things devices. *Softw. Pract. Experience* **51**(3), 595–606 (2021)
30. Peng, G., Zhang, H., Xu, D.: Image emotion analysis based on the distance relation of emotion categories via deep metric learning. In: Magnenat-Thalmann, N., et al. (eds.) *CGI 2021. LNCS*, vol. 13002, pp. 535–547. Springer, Cham (2021). [https://doi.org/10.1007/978-3-030-89029-2\\_41](https://doi.org/10.1007/978-3-030-89029-2_41)
31. Rai Jain, P., Quadri, S., Lalit, M.: Recent trends in artificial intelligence for emotion detection using facial image analysis. In: 2021 Thirteenth International Conference on Contemporary Computing (IC3-2021), pp. 18–36 (2021)
32. Razak, A., Abidin, M., Komiya, R.: Emotion pitch variation analysis in Malay and English voice samples. In: 9th Asia-Pacific Conference on Communications (IEEE Cat. No. 03EX732), vol. 1, pp. 108–112 (2003)
33. Garcia, S., Moreno, J., Fanals, L.: Emotion recognition based on parameterized voice signal analysis [reconocimiento de emociones basado en el análisis de la señal de voz parametrizada]. In: *Actas da 1a Conferência Ibérica de Sistemas e Tecnologias de Informação, Ofir, Portugal, 21 a 23 de Junho de 2006*, vol. 2, pp. 837–854 (2006)
34. Iraki, F., et al.: The expression and recognition of emotions in the voice across five nations: a lens model analysis based on acoustic features. *J. Person. Soc. Psychol.* **111**(5), 686–705 (2016)
35. Mitsuyoshi, S., et al.: Emotion voice analysis system connected to the human brain. In: 2007 International Conference on Natural Language Processing and Knowledge Engineering, pp. 476–484 (2007)
36. Farooque, M., Munoz-Hernandez, S.: Easy fuzzy tool for emotion recognition: prototype from voice speech analysis. In: *IJCCI*, pp. 85–88 (2009)
37. Chaturvedi, I., Satapathy, R., Cavallari, S., Cambria, E.: Fuzzy commonsense reasoning for multimodal sentiment analysis. *Pattern Recogn. Lett.* **125**, 264–270 (2019)
38. Gao, Y., Ohno, Y., Qian, F., Hu, Z., Wang, Z.: The discriminant analysis of the voice expression of emotion - focus on the nursing experience - focus o. In: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 1262–1265 (2013)
39. Manasa, C., Dheeraj, D., Deepth, V.: Statistical analysis of voice based emotion recognition using similarity measures. In: 2019 1st International Conference on Advanced Technologies in Intelligent Control, Environment, Computing & Communication Engineering (ICATIECE), pp. 46–50 (2019)
40. Busso, C., Lee, S., Narayanan, S.: Analysis of emotionally salient aspects of fundamental frequency for emotion detection. *IEEE Trans. Audio Speech Lang. Process.* **17**(4), 582–596 (2009)
41. Hekiert, D., Igras-Cybulska, M.: Capturing emotions in voice: a comparative analysis of methodologies in psychology and digital signal processing. *Ann. Psychol.* **22**(1), 15–34 (2019)

42. Ringeval, F., et al.: Automatic analysis of typical and atypical encoding of spontaneous emotion in the voice of children. In: Proceedings INTERSPEECH 2016, 17th Annual Conference of the International Speech Communication Association (ISCA), 08–12 September 2016, pp. 1210–1214 (2016)
43. Iliev, A., Stanchev, P.: Smart multifunctional digital content ecosystem using emotion analysis of voice. In: Proceedings of the 18th International Conference on Computer Systems and Technologies, vol. Part F132086, pp. 58–64 (2017)
44. Panigrahi, S., Palo, H.: Analysis and recognition of emotions from voice samples using ant colony optimization algorithm. *Lect. Notes Electr. Eng.* **814**, 219–231 (2022)
45. Develasco, M., Justo, R., Zorrilla, A., Inés Torres, M.: Automatic analysis of emotions from the voices/speech in Spanish tv debates. *Acta Polytech. Hung.* **19**(5), 149–171 (2022)
46. Koren, L., Stipančič, T.: Multimodal emotion analysis based on acoustic and linguistic features of the voice. In: Meiselwitz, G. (ed.) HCII 2021. LNCS, vol. 12774, pp. 301–311. Springer, Cham (2021). <https://doi.org/10.1007/978-3-030-77626-8-20>
47. Sukumaran, P., Govardhanan, K.: Towards voice based prediction and analysis of emotions in ASD children. *J. Intell. Fuzzy Syst.* **41**(5), 5317–5326 (2021)
48. Chengeta, K.: Comparative analysis of emotion detection from facial expressions and voice using local binary patterns and Markov models: computer vision and facial recognition. In: Proceedings of the 2nd International Conference on Vision, Image and Signal Processing (2018)
49. Riley, P.: Three pitfalls to avoid in machine learning. *Nature* **572**, 27–29 (2019)
50. Eyben, F., et al.: The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. *IEEE Trans. Affect. Comput.* **7**(2), 190–202 (2015)
51. Wujek, B., Hall, P., Günes, F.: Best practices for machine learning applications. SAS Institute Inc (2016)
52. Biderman, S., Scheirer, W.J.: Pitfalls in machine learning research: reexamining the development cycle (2020). [arXiv:2011.02832](https://arxiv.org/abs/2011.02832)
53. Shannon, C.E.: A mathematical theory of communication. *Bell Syst. Tech. J.* **27**(3), 379–423 (1948)
54. Moukadem, A., Dieterlen, A., Brandt, C.: Shannon entropy based on the s-transform spectrogram applied on the classification of heart sounds. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 704–708 (2013)
55. Wang, X.P., Liu, C.C., Li, Y.Y., Sun, C.R.: Heart sound segmentation algorithm based on high-order Shannon entropy. *Jilin Daxue Xuebao (Gongxueban) J. Jilin Univ. (Eng. Technol. Ed.)* **40**(5), 1433–1437 (2010)
56. Fourier, J.B.J., Darboux, G., et al.: *Théorie analytique de la chaleur*, vol. 504. Didot Paris (1822)
57. Proakis, Jonh G., D.G.M.: *Tratamiento Digital de Señales*. Prentice Hall, Madrid (2007)
58. Fadlelmoula, A., Pinho, D., Carvalho, V., Catarino, S., Minas, G.: Fourier transform infrared (FTIR) spectroscopy to analyse human blood over the last 20 years: a review towards lab-on-a-chip devices. *Micromachines* **13**(2), 187 (2022)
59. Gómez-Echavarría, A., Ugarte, J., Tobón, C.: The fractional Fourier transform as a biomedical signal and image processing tool: a review. *Biocybernetics Biomed. Eng.* **40**(3), 1081–1093 (2020)

60. Shakya, B., Shrestha, P., Teppo, H.R., Rieppo, L.: The use of Fourier transform infrared (FTIR) spectroscopy in skin cancer research: a systematic review. *Appl. Spectrosc. Rev.* **56**(5), 1–33 (2020)
61. Su, K.Y., Lee, W.L.: Fourier transform infrared spectroscopy as a cancer screening and diagnostic tool: a review and prospects. *Cancers* **12**(1), 115 (2020)
62. Hertzog, J., Mase, C., Hubert-Roux, M., Afonso, C., Giusti, P., Barrère-Mangote, C.: Characterization of heavy products from lignocellulosic biomass pyrolysis by chromatography and Fourier transform mass spectrometry: a review. *Energy Fuels* **35**(22), 17979–18007 (2021)
63. Giechaskiel, B., Clairotte, M.: Fourier transform infrared (FTIR) spectroscopy for measurements of vehicle exhaust emissions: a review. *Appl. Sci. (Switzerland)* **11**(16), 7416 (2021)
64. Bahureksa, W., et al.: Soil organic matter characterization by Fourier transform ion cyclotron resonance mass spectrometry (FTICR MS): a critical review of sample preparation, analysis, and data interpretation. *Environ. Sci. Technol.* **55**(14), 9637–9656 (2021)
65. Zhang, X., et al.: Application of Fourier transform ion cyclotron resonance mass spectrometry in deciphering molecular composition of soil organic matter: a review. *Sci. Total Environ.* **756**, 144140 (2021)
66. He, Z., Liu, Y.: Fourier transform infrared spectroscopic analysis in applied cotton fiber and cottonseed research: a review. *J. Cotton Sci.* **25**(2), 167–183 (2021)
67. Chirman, D., Pleshko, N.: Characterization of bacterial biofilm infections with Fourier transform infrared spectroscopy: a review. *Appl. Spectrosc. Rev.* **56**(8–10), 673–701 (2021)
68. Veerasingam, S., et al.: Contributions of Fourier transform infrared spectroscopy in microplastic pollution research: a review. *Crit. Rev. Environ. Sci. Technol.* **51**(22), 2681–2743 (2021)
69. Hirschmugl, C., Gough, K.: Fourier transform infrared spectrochemical imaging: Review of design and applications with a focal plane array and multiple beam synchrotron radiation source. *Appl. Spectrosc.* **66**(5), 475–491 (2012)
70. Oane, M., Mahmood, M., Popescu, A.: A state-of-the-art review on integral transform technique in laser-material interaction: Fourier and non-Fourier heat equations. *Materials* **14**(16), 4733 (2021)
71. John, A., Khanna, K., Prasad, R., Pillai, L.: A review on application of Fourier transform in image restoration. In: 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), pp. 389–397 (2020)
72. Ghani, H., Malek, M., Azmi, M., Muril, M., Azizan, A.: A review on sparse fast Fourier transform applications in image processing. *Int. J. Electr. Comput. Eng.* **10**(2), 1346–1351 (2020)
73. Su, W.H., Sun, D.W.: Fourier transform infrared and Raman and hyperspectral imaging techniques for quality determinations of powdery foods: a review. *Compr. Rev. Food Sci. Food Saf.* **17**(1), 104–122 (2018)
74. Pralle, R., White, H.: Symposium review: big data, big predictions: utilizing milk Fourier-transform infrared and genomics to improve hyperketonemia management. *J. Dairy Sci.* **103**(4), 3867–3873 (2020)
75. Chai, J., et al.: Review of mems based Fourier transform spectrometers. *Micro-machines* **11**(2), 1–28 (2020)
76. Valand, R., Tanna, S., Lawson, G., Bengtström, L.: A review of Fourier transform infrared (FTIR) spectroscopy used in food adulteration and authenticity investigations. *Food Addit. Contam. Part A Chem. Anal. Control Exposure Risk Assess.* **37**(1), 19–38 (2020)

77. Bureau, S., Cozzolino, D., Clark, C.: Contributions of Fourier-transform mid infrared (FT-MIR) spectroscopy to the study of fruit and vegetables: a review. *Postharvest Biol. Technol.* **148**, 1–14 (2019)
78. Takahama, S., et al.: Atmospheric particulate matter characterization by Fourier transform infrared spectroscopy: a review of statistical calibration strategies for carbonaceous aerosol quantification in us measurement networks. *Atmos. Meas. Tech.* **12**(1), 525–567 (2019)
79. Lin, H.C., Ye, Y.C.: Reviews of bearing vibration measurement using fast Fourier transform and enhanced fast Fourier transform algorithms. *Adv. Mech. Eng.* **11**(1), 168781401881675 (2019)
80. Ravisankar, M., Sreenivas, A.: A review on estimation of chromatic dispersion using fractional Fourier transform in optical fiber communication. In: 2018 International Conference on Smart Systems and Inventive Technology (ICSSIT), pp. 223–228 (2018)
81. Calabró, E., Magazú, S.: A review of advances in the analysis of biological systems by means of fourier transform infrared (FTIR) spectroscopy. In: Moore, E. (ed.) *Fourier Transform Infrared Spectroscopy (FTIR): Methods, Analysis and Research Insights*, pp. 1–32. Nova Science Publishers Inc (2016)
82. Chen, Y., Zou, C., Mastalerz, M., Hu, S., Gasaway, C., Tao, X.: Applications of micro-Fourier transform infrared spectroscopy (FTIR) in the geological sciences—a review. *Int. J. Mole. Sci.* **16**(12), 30223–30250 (2015)
83. Kaushik, C., Gautam, T., Elamaram, V.: A tutorial review on discrete Fourier transform with data compression application. In: 2014 International Conference on Green Computing Communication and Electrical Engineering (ICGCCCE) (2014)
84. Hauchecorne, B., Lenaerts, S.: Unravelling the mysteries of gas phase photocatalytic reaction pathways by studying the catalyst surface: a literature review of different Fourier transform infrared spectroscopic reaction cells used in the field. *J. Photochem. Photobiol. C Photochem. Rev.* **14**(1), 72–85 (2013)
85. Zappa, E., Busca, G.: Static and dynamic features of Fourier transform profilometry: a review. *Opt. Lasers Eng.* **50**(8), 1140–1151 (2012)
86. Xiong, J., Zhang, Z., Chen, X.: Multidimensional entropy evaluation of non-uniform distribution of assembly features in precision instruments. *Precision Eng.* **77**, 1–15 (2022)
87. Sütçü, M.: Disutility entropy in multi-attribute utility analysis. *Comput. Ind. Eng.* **169**, 108189 (2022)
88. You, J., Yu, C., Sun, J., Chen, J.: Generalized maximum entropy based identification of graphical arma models. *Automatica* **141**, 110319 (2022)
89. Wang, Y., Wang, Z., Chen, L., Gu, J.: Experimental study on critical characteristics of self-organized behavior of concrete under uniaxial compression based on AE characteristic parameters information entropy. *J. Mater. Civil Eng.* **34**(7) (2022)
90. Yuan, Z., Chen, H., Li, T.: Exploring interactive attribute reduction via fuzzy complementary entropy for unlabeled mixed data. *Pattern Recognit.* **127**, 108651 (2022)
91. Zhang, X., Zhou, Y., Tang, X., Fan, Y.: Three-way improved neighborhood entropies based on three-level granular structures. *Int. J. Mach. Learn. Cybernetics* **13**(7), 1861–1890 (2022)
92. Kadian, R., Kumar, S.: New fuzzy mean codeword length and similarity measure. *Granular Comput.* **7**(3), 461–478 (2022)

93. Grassl, M., Huber, F., Winter, A.: Entropic proofs of singleton bounds for quantum error-correcting codes. *IEEE Trans. Inform. Theo.* **68**(6), 3942–3950 (2022)
94. Dai, J., Liu, Q.: Semi-supervised attribute reduction for interval data based on misclassification cost. *Int. J. Mach. Learn. Cybernetics* **13**(6), 1739–1750 (2022)
95. Gao, X., Pan, L., Deng, Y.: Cross entropy of mass function and its application in similarity measure. *Appl. Intell.* **52**(8), 8337–8350 (2022)
96. Lo, Y., Hsiao, Y.T., Chang, F.C.: Use electroencephalogram entropy as an indicator to detect stress-induced sleep alteration. *Appl. Sci. (Switzerland)* **12**(10), 4812 (2022)
97. Chirikjian, G.: Entropy, symmetry, and the difficulty of self-replication. *Artif. Life Robot.* **27**(2), 181–195 (2022)
98. Wahdan, M., Altnkaya, M.: Maximum average entropy-based quantization of local observations for distributed detection. *Digit. Sign. Process. A Rev. J.* **123**, 103427 (2022)
99. Livingstone, S.R., Russo, F.A.: The Ryerson audio-visual database of emotional speech and song (RAVDESS): a dynamic, multimodal set of facial and vocal expressions in north American English. *PloS one* **13**(5), e0196391 (2018)
100. McFee, B., et al.: librosa: Audio and music signal analysis in python. In: *Proceedings of the 14th python in science conference*, vol. 8, pp. 18–25. Citeseer (2015)
101. The SciPy community: Scipy (2022) Accessed 10 April 2022
102. The SciPy community: Scipyfft (2022) Accessed 10 April 2022
103. Cover, T., Thomas, J.: *Elements of Information Theory*. Wiley (2012)
104. Veena, S., Aravindhar, D.J.: Sound classification system using deep neural networks for hearing impaired people. *Wireless Pers. Commun.* **126**, 385–399 (2022). <https://doi.org/10.1007/s11277-022-09750-7>
105. Khamparia, A., Gupta, D., Nguyen, N.G., Khanna, A., Pandey, B., Tiwari, P.: Sound classification using convolutional neural network and tensor deep stacking network. *IEEE Access* **7**, 7717–7727 (2019)
106. Kwon, S.: A cnn-assisted enhanced audio signal processing for speech emotion recognition. *Sensors* **20**(1), 183 (2019)
107. Xia, X., Pan, J., Wang, Y.: Audio sound determination using feature space attention based convolution recurrent neural network. In: *ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3382–3386 (2020)
108. Shannon, C.E.: Prediction and entropy of printed english. *Bell Syst. Tech. J.* **30**(1), 50–64 (1951)
109. Chamishka, S., et al.: A voice-based real-time emotion detection technique using recurrent neural network empowered feature modelling. *Multimed. Tools Appl.* 1–22 (2022). <https://doi.org/10.1007/s11042-022-13363-4>
110. Lee, M.C., Yeh, S.C., Chang, J.W., Chen, Z.Y.: Research on Chinese speech emotion recognition based on deep neural network and acoustic features. *Sensors* **22**(13), 4744 (2022)