Alexey Karpov
Vlado Delić (Eds.)

LNAI 15300

Speech and Computer

26th International Conference, SPECOM 2024
Belgrade, Serbia, November 25–28, 2024
Proceedings, Part II

2 Part II

Springer

Lecture Notes in Computer Science

Lecture Notes in Artificial Intelligence 15300

Founding Editor

Jörg Siekmann

Series Editors

The series Lecture Notes in Artificial Intelligence (LNAI) was established in 1988 as a topical subseries of LNCS devoted to artificial intelligence.

The series publishes state-of-the-art research results at a high level. As with the LNCS mother series, the mission of the series is to serve the international R & D community by providing an invaluable service, mainly focused on the publication of conference and workshop proceedings and postproceedings.

Alexey Karpov · Vlado Delić
Editors

Speech and Computer

26th International Conference, SPECOM 2024
Belgrade, Serbia, November 25–28, 2024
Proceedings, Part II

Springer

Editors
Alexey Karpov
St. Petresburg Federal Research Center
of the Russian Academy of Sciences
St. Petresburg, Russia

Vlado Delić
University of Novi Sad
Novi Sad, Serbia

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Artificial Intelligence
ISBN 978-3-031-78013-4 ISBN 978-3-031-78014-1 (eBook)
https://doi.org/10.1007/978-3-031-78014-1

LNCS Sublibrary: SL7 – Artificial Intelligence

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

SPECOM 2024 Preface

SPECOM is a conference with a long tradition that attracts researchers in the area of speech technology, including automatic speech recognition and understanding, text–to–speech synthesis, speaker and language recognition, as well as related domains like digital speech processing, natural language processing, text analysis, computational paralinguistics, multi–modal speech, and data processing or human–computer interaction. The SPECOM conference is an ideal platform for know–how exchange – especially for experts working on Slavic languages (e.g. Russian, Serbian, Croatian, Polish, Bulgarian, Czech, etc.) or other inflectional spoken languages – including both under–resourced and regular well–resourced ones.

The International Conference on Speech and Computer (SPECOM) has become a regular event since the first SPECOM, held in St. Petresburg, Russia, in October 1996. The SPECOM conference series was established more than 28 years ago by the St. Petresburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS).

In its long history, the SPECOM conference was organized alternately by the St. Petresburg Federal Research Center of the Russian Academy of Sciences (SPC RAS)/SPIIRAS and by the Moscow State Linguistic University (MSLU) in their home towns. Furthermore, in 1997 it was organized by the Cluj–Napoca subsidiary of the Research Institute for Computer Technique (Romania), in 2005 and 2015 by the University of Patras (in Patras and Athens, Greece), in 2011 by the Kazan Federal University (in Kazan, Russia), in 2013 by the University of West Bohemia (in Pilsen, Czech Republic), in 2014 by the University of Novi Sad (in Novi Sad, Serbia), in 2016 by the Budapest University of Technology and Economics (in Budapest, Hungary), in 2017 by the University of Hertfordshire (in Hatfield, UK), in 2018 by the Leipzig University of Telecommunications (in Leipzig, Germany), in 2019 by the Bogaziçi University (in Istanbul, Turkey), in 2020 and 2021 by SPC RAS/SPIIRAS (fully online), in 2022 by the KIIT (in Gurugram, New Delhi, India), and in 2023 by the IIT/IIIT Dharwad (in Hubli–Dharwad, Karnataka, India).

SPECOM 2024 was the 26th event in the conference series (https://specom2024.ftn.uns.ac.rs), and the second time SPECOM was in the Republic of Serbia. SPECOM 2024 was organized jointly by the Faculty of Technical Sciences at the University of Novi Sad and the School of Electrical Engineering at the University of Belgrade in cooperation with the Telecommunications Society of Serbia. The conference was held between the 25th and 28th November 2024, in a hybrid format, mostly in–person in the capital of Serbia, Belgrade, at the Crowne Plaza Hotel and online via video conferencing. Moreover, SPECOM 2024 was organized jointly and in parallel with the 32nd Telecommunications Forum TELFOR 2024 (https://www.telfor.rs/en). SPECOM 2024 was sponsored and supported by the Science Fund of the Republic of Serbia, as well as by the International Speech Communication Association (ISCA).

During SPECOM 2024, two keynote lectures were given by Dr.–Ing. Kraljevski (Fraunhofer Institute for Ceramic Technologies and Systems IKTS, Dresden, Germany) jointly with his German colleagues on "Preserving Language Heritage Through Speech Technology: The Case of Upper Sorbian", as well as by Prof. Milan Sečujski jointly with his colleagues from the Faculty of Technical Sciences, University of Novi Sad and AlfaNum company, Novi Sad, Serbia on "Retrospective and Perspectives of TTS & STT Technology Development and Implementation for South Slavic Under-Resourced Languages".

This volume contains a collection of submitted papers presented at SPECOM 2024, which were thoroughly reviewed by members of the Program Committee and additional reviewers consisting of over 80 experts in the conference topic areas. In total, 53 regular full papers out of 90 submissions to SPECOM 2024 were carefully selected by the Program Committee members for oral presentation at the conference, as well as for inclusion in these SPECOM 2024 proceedings. Theoretical and more general contributions were presented in common plenary sessions. Problem–oriented sessions as well as panel discussions brought together specialists in niche problem areas with the aim of exchanging knowledge and skills resulting from research projects of all kinds.

We would like to express our gratitude to all authors for providing their papers on time, to the members of the SPECOM 2024 Program Committee for their careful reviews and paper selection, and to the editors and correctors for their hard work in preparing the conference proceedings. Special thanks are due to the members of the SPECOM 2024 Organizing Committee for their tireless effort and enthusiasm during the conference organization. We are also grateful to the Faculty of Technical Sciences at the University of Novi Sad, the School of Electrical Engineering at the University of Belgrade, and the Telecommunications Society of Serbia for organizing and hosting the 26th International Conference on Speech and Computer, SPECOM 2024, in Belgrade.

November 2024

Alexey Karpov
Vlado Delić

Organization

General Chairs

Vlado Delić	University of Novi Sad, Serbia
Alexey Karpov	St. Petresburg Federal Research Center of the Russian Academy of Sciences, Russia

Program Committee

Alexey Karpov (Chair)	St. Petresburg Federal Research Center of the Russian Academy of Sciences, Russia
Vlado Delić (Chair)	University of Novi Sad, Serbia
Shyam Agrawal	KIIT Gurugram, India
Jahangir Alam	Computer Research Institute of Montreal, Canada
Shahin Amiriparian	Technical University of Munich, Germany
Elias Azarov	Belarusian State University of Informatics and Radioelectronics, Belarus
Milana Bojanić	University of Novi Sad, Serbia
Nick Campbell	Trinity College Dublin, Ireland
Vladimir Chuchupal	Federal Research Center "Computer Science and Control" of Russian Academy of Sciences, Russia
Andrea Corradini	Design School Kolding, Denmark
Govind D.	K L University, India
Olivier Deroo	Acapela Group, Belgium
Denis Dresvyanskiy	Ulm University, Germany
Anna Esposito	Università degli Studi della Campania "Luigi Vanvitelli", Italy
Vera Evdokimova	Saint-Petersburg State University, Russia
Nikos Fakotakis	University of Patras, Greece
Mauro Falcone	Fondazione Ugo Bordoni, Italy
Abderrahim Fathan	Computer Research Institute of Montreal, Canada
Olga Frolova	Saint-Petersburg State University, Russia
Jovan Galić	University of Banja Luka, Bosnia and Herzegovina
Suryakanth Gangashetty	KLEF, India
Philip N. Garner	IDIAP Research Institute, Switzerland

Branislav Gerazov	Saints Cyril and Methodius University in Skopje, North Macedonia
Barbara Gili Fivela	Università del Salento, Italy
Gábor Gosztolya	University of Szeged, Hungary
Ivan Gruber	University of West Bohemia, Czech Republic
Denis Ivanko	St. Petresburg Federal Research Center of the Russian Academy of Sciences, Russia
Nikša Jakovljević	University of Novi Sad, Serbia
Ildar Kagirov	St. Petresburg Federal Research Center of the Russian Academy of Sciences, Russia
Alexey Kashevnik	St. Petresburg Federal Research Center of the Russian Academy of Sciences, Russia
Heysem Kaya	Utrecht University, The Netherlands
Maria Khokhlova	Saint-Petersburg State University, Russia
Irina Kipyatkova	St. Petresburg Federal Research Center of the Russian Academy of Sciences, Russia
Liliya Komalova	Moscow State Linguistic University, Russia
Evgeny Kostyuchenko	Tomsk State University of Control Systems and Radioelectronics, Russia
Yanxiong Li	South China University of Technology, China
Natalia Loukachevitch	Moscow State University, Russia
Elena Lyakso	Saint-Petersburg State University, Russia
Ilya Makarov	Artificial Intelligence Research Institute, Russia
Olesia Makhnytkina	ITMO University, Russia
Maxim Markitantov	St. Petresburg Federal Research Center of the Russian Academy of Sciences, Russia
Konstantin Markov	University of Aizu, Japan
Yuri Matveev	ITMO University, Russia
Peter Mihajlik	Budapest University of Technology and Economics, Hungary
Nikolay Mikhaylovskiy	Tomsk State University, Russia
Bernd Möbius	Saarland University, Germany
Sebastian Möller	Technical University Berlin, Germany
Ruban Nersisson	Vellore Institute of Technology University, India
Aleksandar Nešković	University of Belgrade, Serbia
Tijana Nosek	University of Novi Sad, Serbia
Dariya Novokhrestova	Tomsk State University of Control Systems and Radioelectronics, Russia
Sergey Novoselov	STC-Innovations Ltd., Russia
Nick A. Petrovsky	Belarusian State University of Informatics and Radioelectronics, Belarus
Lidia Pivovarova	University of Helsinki, Finland
Branislav Popović	University of Novi Sad, Serbia

Vsevolod Potapov	Lomonosov Moscow State University, Russia
Rodmonga Potapova	Moscow State Linguistic University, Russia
Sergey Rybin	ITMO University, Russia
Dmitry Ryumin	St. Petresburg Federal Research Center of the Russian Academy of Sciences, Russia
Elena Ryumina	St. Petresburg Federal Research Center of the Russian Academy of Sciences, Russia
Milan Sečujski	University of Novi Sad, Serbia
Tatiana Sherstinova	HSE University, St. Petresburg, Russia
Nickolay Shmyrev	Alpha Cephei Inc., Russia
Vasiliki Simaki	Lancaster University, UK
Nikola Simić	University of Novi Sad, Serbia
Pavel Skrelin	Saint-Petersburg State University, Russia
Tatiana Sokoreva	Moscow State Linguistic University, Russia
Victor Sorokin	Institute for Information Transmission Problems of the Russian Academy of Sciences, Moscow, Russia
Ajay Srinivasamurthy	Amazon Alexa, India
Siniša Suzić	University of Novi Sad, Serbia
Jianhua Tao	Institute of Automation, Chinese Academy of Sciences, China
Ivan Tashev	Microsoft, USA
Natalia Tomashenko	University of Avignon, France
Laszlo Toth	University of Szeged, Hungary
Isabel Trancoso	INESC-ID/IST, University of Lisbon, Portugal
Jan Trmal	Johns Hopkins University, USA
Liliya Tsirulnik	Stenograph LLC, USA
Alena Velichko	St. Petresburg Federal Research Center of the Russian Academy of Sciences, Russia
Vasilisa Verkhodanova	University of Groningen, Campus Fryslan, The Netherlands
Zeynep Yucel	Okayama University, Japan
Miloš Železný	University of West Bohemia, Czech Republic
Jerneja Žganec Gros	Alpineon, Slovenia

Additional Reviewers

Nikolay Bobrov
Lidija Krstanović
Bin Liu
Danila Mamontov
Yong Ren
Vuk Stanojev
Anton Stepikhov

Organizing Committee

Vlado Delić (Chair)	University of Novi Sad, Serbia
Milan Sečujski	University of Novi Sad, Serbia
Branislav Popović	University of Novi Sad, Serbia
Milana Bojanić	University of Novi Sad, Serbia
Nikola Simić	University of Novi Sad, Serbia
Nikša Jakovljević	University of Novi Sad, Serbia
Siniša Suzić	University of Novi Sad, Serbia
Tijana Nosek	University of Novi Sad, Serbia
Vuk Stanojev	University of Novi Sad, Serbia
Mladen Koprivica	University of Belgrade, Serbia
Jelena Ćertić	University of Belgrade, Serbia
Alexey Karpov	SPC RAS, Russia
Dmitry Ryumin	SPC RAS, Russia
Irina Kipyatkova	SPC RAS, Russia
Ildar Kagirov	SPC RAS, Russia
Alexandr Axyonov	SPC RAS, Russia

Contents – Part II

Digital Speech Processing

Natural Language Processing

Contents – Part I

Automatic Detection of Irony Based on Acoustic Features and Facial Expressions

Uliana Kochetkova(✉), Pavel Skrelin, Vera Evdokimova, Nikolai Borisov, Pavel Scherbakov, Petr Fedkin, and Rada German

Saint Petersburg State University, 7-9 Universitetskaya Embankment, St. Petersburg, Russia
{u.kochetkova,p.skrelin,v.evdokimova,n.borisov,p.scherbakov,
p.fedkin}@spbu.ru

Abstract. The current study deals with the automatic analysis of verbal irony using artificial neural networks. Detection of verbal irony is an important task nowadays, because the effectiveness of the communication depends on the correct interpretation of sentences with an ambiguous meaning. In the case, when the context is lacking, the correct sense can be understood not from the lexical content, but through phonetic features, as well as through co-speech mimics and gestures. Thus we accomplished a new research on the material of the multimedia corpus of Russian ironic speech, which contains the detailed phonetic annotation and irony evaluation by native listeners in perceptual auditory experiments. Two types of automated analysis were accomplished: based on acoustic feature and facial expression extraction. The use of the fully connected neural network and of the Wav2Vec 2.0 model for the automatic irony detection in audio signal demonstrated high level of irony recognition. We also tested on a part of the corpus the recognition of ironic facial expressions in video signal using convolutional neural network and the PyFeat library, which allowed us to conclude that this model can give good results when we increase the amount of the material.

Keywords: Irony · Multimedia Speech Corpus · Artificial Neural Networks · Acoustic Feature Extraction · Facial Expression Analysis

1 Introduction

Nowadays the volume of information is rapidly growing, the task of effectively understanding the meaning of statements becomes urgent and complex. One of the challenges in this context is the recognition of irony – a phenomenon that carries polysemy and ambiguity. Irony is often confusing and difficult to understand even for humans, which poses a major obstacle to modern speech recognition technologies [4, 10, 15, 18].

In this context, the use of neural network technologies represents a promising way to solve the problem of recognizing any kind of acoustic and paralinguistic information that influences the speech and gives the information about the correct way to recognize the text of the utterance. Neural networks demonstrate remarkable ability to analyze complex patterns in data. The development of deep learning methods makes it possible

A. Karpov and V. Delić (Eds.): SPECOM 2024, LNAI 15300, pp. 70–82, 2025.
https://doi.org/10.1007/978-3-031-78014-1_6

to effectively extract features from both texts and audio, which is becoming an important element in the development of irony recognition systems. That is why it is important to test the performance of the machine learning systems and compare it to the information obtained by expert acoustic analysis and perceptual evaluation of ironic or non-ironic meaning by native listeners.

The relevance of this study is enhanced in the context of the increased influence of irony in online environments, where it is actively used in digital dialogues and sociocultural expressions. The need for automated irony recognition tools becomes an important task in the context of various fields, such as sentiment analysis, cybersecurity, reputation management and many others, where it is important to correctly interpret and analyze the emotional and semantic context of statements.

Thus, research on irony recognition using neural network technologies is becoming important in light of modern requirements for information processing and automated analysis of audio data. However, the analysis of text data using deep learning methods has been much better developed compared to the audio data [35]. This analysis is commonly based on the messages in social media and networks, as well as on comments on online-newspapers [2, 6, 11, 12, 16, 19, 20, 26, 27].

Although the acoustic analysis of ironic speech has been carried out [7, 17, 30], irony detection using neural network technologies is still not a fully solved task. One of the most important investigation in this field has been recently done [14] on the material of English naturalistic conversational speech with the extraction of acoustic features including various parameters of pitch, intensity, timing, voice quality and Mel-Frequency Cepstral Coefficients (MFCC). The data obtained showed a high accuracy of irony detection in the material.

Research on facial expressions using deep learning has mostly dealt with various emotions [31], while irony was studied with eye-tracking methods [28]. At the same time, the instruments of facial expression analysis in video signal are well elaborated nowadays [3, 5, 8, 9, 13, 24, 25, 29, 32, 33].

In order to test the ANN for the automatic irony detection in Russian speech we used the Multimedia corpus of Russian ironic speech described in [22, 23, 34].

2 Multimedia Corpus of Russian Ironic Speech

Reading Material and Experimental Design. We recorded the material in an equipped studio at the Department of Phonetics of Saint Petersburg State University. Native Russian speakers read the sets of short texts and long coherent texts. The texts were printed on paper. One session of recordings had 40 min duration as maximum. The speakers read one set of short texts (2–4 sentences long) and one coherent text. They were given a prompt to read in the way they would pronounce such sentences in their everyday life. The term "irony" was avoided in the reading material. The reader was supposed to read a sentence with or without irony depending on the surrounding context. The context itself was read as well. In order to help the speaker, we constructed short ironic and non-ironic monologues and dialogues with homonymous target fragments. The recorded material has been presented in 2 formats: one folder with whole contexts with the target fragments inside and another folder with target fragments fully annotated in Praat (containing detailed phonetic annotation, as well as native speaker's evaluation of presence

or absence of irony in the target fragment). The annotation is presented in TextGrid files with the following levels:

- target phrase – level s,
- stressed vowel – level v,
- stressed syllable – level syl,
- context – level c.

The contexts folder contains the audio of the entire utterance with the appropriate markup.

Perceptual Analysis. The target fragments extracted from the ironic and non-ironic short texts were presented to listeners. There were no context or lexical marker of presence or absence of the ironic meaning. The listeners were suggested to choose the corresponding written context for the audio fragment they heard.

Acoustic Analysis. The expert acoustic analysis of the target fragments evaluated by speakers as ironic and non-ironic allowed finding salient perceptually relevant features of irony. These features were of 2 types: increasing of values (stress vowel duration, intensity level, melodic range, spectral density) or decreasing of the same values. The concrete strategy depended on the sentence type and individual characteristics of a speaker. But the most important was the contrast between ironic and non-ironic homonymous sentences that was present in all speech material.

Audio Signal Modifications. After establishing acoustic cues of irony we studied the role each of them plays in a complex acoustic signal. For this purpose we carried out the experiments with modified stimuli. We modified the duration, the intensity level and the melodic pattern of the original non-ironic stimuli to make them ironic, and vice-versa: we turned ironic statements into neutral ones. There were isolated modifications of each parameter and complex modifications. The results of the perceptual analysis showed that the melodic parameter change is necessary for a successful evaluation of the presence or absence of irony. Also it was showed that the modification of ironic fragments into non-ironic were less effective than turning non-ironic into ironic. Such result was due to the voice quality, which was not modified.

Expert Analysis of Paralinguistic Cues. At this stage of the analysis we compared gestures and mimics in ironic and non-ironic fragments using ELAN software. The data obtained showed the difference in types of gestures and facial expressions in two types of fragments. But a more precise synchronization of the video signal with the annotated audio signal is possible only using the mathematical models of analysis, which was made in the current study.

3 Automatic Irony Detection Based on Acoustic Features

The task of this part of the work was to develop and implement a method for recognizing irony in speech using neural network technologies. The main focus was on creating an effective acoustic model capable of automatically identifying ironic statements, which is of great practical importance for improving the quality of speech data analysis.

The object of this study is the process of recognizing irony in speech using neural network technologies using the example of a corpus of ironic speech. How it can be improved through preliminary acoustic analysis of the material.

The subject of this part of the study was to consider various acoustic characteristics of irony and methods for their isolation, used in the field of neural network technologies. We analysed the approaches and neural network architectures, which can be effective in solving this problem, as well as what features of an ironic statement should be taken into account when developing appropriate models.

The following tasks were accomplished:

1. Collecting and preparing data for training and additional training of irony recognition models.
2. Assessing the effectiveness of existing speech recognition systems.
3. Studying and testing the Wav2Vec 2.0 model.
4. Collecting a dataset with the acoustic characteristics of the speech signal.
5. Training a fully connected irony recognition model.
6. Retraining the Wav2Vec 2.0 model to recognize irony.
7. Identifying the most effective algorithm for recognizing irony.

3.1 Material

For this part of the study we used recordings of 56 speakers (32 women and 24 men); we selected only the target fragments (with no surrounding context) from 4499 audio files with corresponding TextGrid files containing annotation in Praat Software. We didn't add the contexts, because only the target fragments contain the acoustic information, which is necessary for the acoustic feature extraction in a dataset.

3.2 Methods

We conducted two experiments:

- Training a fully connected neural network based on a data frame with acoustic characteristics.
- Additional training of the Wav2Vec 2.0 model based on audio files.

The first part of the work was to extract relevant acoustic sound characteristics that are responsible for irony in speech and subsequent training of a fully connected neural network for the binary irony/neutral classification task. Based on the findings of previous studies, a number of phonetic characteristics were selected to form the training dataset:

- Duration of stressed vowel and syllable;
- Intensity of stressed vowel, syllable and phrase;
- Average pitch frequency;
- Melodic range;
- Speech rate.

To collect these characteristics, the Parselmouth library was chosen. MFCC (Mel-Frequency Cepstral Coefficients) coefficients were also calculated using the LibROSA library, which provides tools for working with audio files in Python and a convenient

interface for extracting MFCC and other acoustic features from audio signals. The resulting dataset is saved in csv format and contains 4499 rows and 24 columns with features.

The pre-trained Russian language model "jonatasgrosman/wav2vec2-large-xlsr-53-russian" was loaded. This is a pre-trained Wav2Vec 2.0 model available in the Hugging Face Transformers library. It is designed for processing and analyzing audio data in Russian.

3.3 Results

The results obtained during training and testing of the model can be seen in Table 1. The resulting model showed 94% accuracy on the training set and 72% on the test set (Table 2).

Table 1. Accuracy on training and test samples before and after Feature importance

Heading level	Accuracy	Accuracy (Feature Importance)
X_train	0.94	0.96
X_test	0.72	0.75

Table 2. Classification report of a trained fully connected neural network

	Precision	Recall	f1-score
Irony	0.73	0.76	0.74
Non-irony	0.76	0.73	0.75
Accuracy	0.75		

Next, it was decided to analyze the usefulness of the features that were transferred to training (Feature Importance) in order to increase the recognition accuracy. One of these methods is to calculate the average value of the absolute values of gradients for features at each learning step. The results are displayed in Fig. 1, where features with higher absolute gradient values will be considered more important. The X-axis on the graph will display the importance value of the features, and the Y-axis will display the corresponding features.

The data obtained show that the following signs were the least indicative: stressed vowel intensity, syllable intensity, melodic range, 1, 11, 12, 13 MFCC coefficients. These characteristics were removed from the training dataset. Then the model was trained again and the accuracy increased to 96% on the training set and to 75% on the test set.

The pre-trained Wav2Vec 2.0 model showed an accuracy of 72% on the test set, which is comparable to the results of a model based on a fully connected neural network. Table 3 shows the classification report.

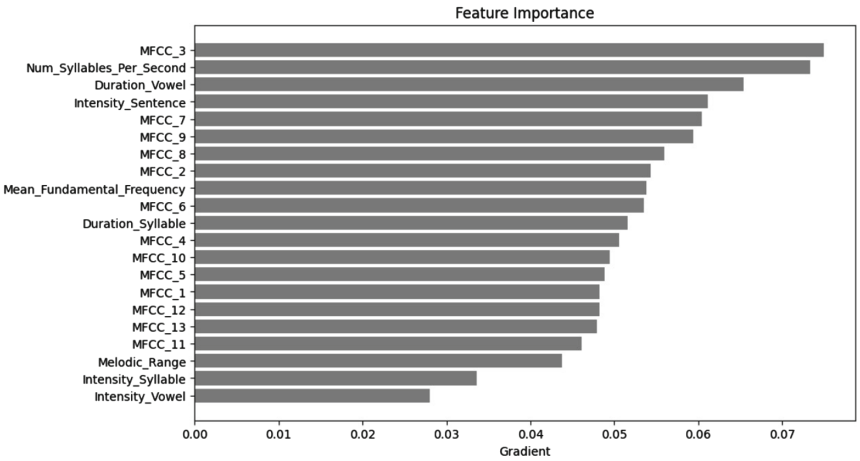

Fig. 1. The results of the analysis of the usefulness of features that were transferred to training (Feature Importance) to increase recognition accuracy. Calculation of the average value of the absolute values of gradients by features at each learning step

Table 3. Classification report for the retrained Wav2Vec 2.0 model

	Precision	Recall	f1-score
Irony	0.79	0.69	0.71
Non-irony	0.75	0.66	0.70
Accuracy	0.73		

To increase the percentage of irony recognition, it was decided to use the data about perceptual evaluation of the corpus. Since when recording the corpus, the speaker could pronounce the phrase more or less ironically, it was impossible not to take into account the percentage of phrases that are not perceived as such, therefore, to increase the efficiency of the algorithm, we used the additional perceptual checking of the target phrases.

Phrases identified as ironic or neutral by more than 60% of respondents were selected for further research. As a result, the original corpus was reduced by 23%. The final corpus of ironic speech included 3,464 audio recordings.

Using the final corpus as an example, two experiments were repeated with training a fully connected neural network and additional training of the Wav2Vec 2.0 model. Table 4 presents the final results of the study.

Table 4. Comparative results of two experiments after perceptual evaluation of the corpus

Accuracy	Before perceptual evaluation of the corpus	After perceptual evaluation of the corpus
Fully connected neural network	0.75	0.82
Wav2Vec 2.0	0.72	0.80

4 Automatic Irony Detection Based on Facial Expressions

The detailed phonetic annotation that was carried out in the Multimedia Corpus of Russian ironic speech allowed us to determine the boundaries of the linguistic units, in which the listeners evaluated the presence or absence of ironic meaning. Thus, we could use the audio files and the corresponding TextGrid files with annotation and synchronize them with the video recordings that were made in parallel with the audio recordings. It let us to find the same linguistic units within the video files and annotate them as ironic or non-ironic. The goal was to test the model capacity to recognize irony and the related emotions using facial feature analysis.

4.1 Material and Method

For this preliminary analysis of the model capacity we selected the recordings of 1 set of short texts read by 9 speakers who differed by their strategies in using facial expressions and gestures when expressing irony in speech. We based our choice on the results of our studies of facial expressions and gestures using the ELAN software that we conducted at previous stage of the corpus analysis [22].

Preparing Data for ANN Training with the PyFeat Library. To train a neural network capable of detecting fragments with irony from video signal, it is required to prepare two types of datasets – training and verification. Based on the training data set, basic training takes place, and a test set is needed to determine that the network is not retrained to recognize only training data, but is able to recognize those that are not involved in training.

As the training and verification datasets should contain the values of changes in facial facial muscles (action units), the PyFeat library was used to obtain this information from preprocessed video clips.

PyFeat provides a Detector class with which it is possible to detect changes in facial muscles and save these values to a separate file. Thus, we created the make_prediction function based on this concept. An example of a generated table based on the created csv file is given below (Fig. 2). Using this method a set of 558 fragments of 20 frames was assembled. Each video clip can be either with the presence of irony or without it.

Informativeness of the Training Data. Before starting to train the network, it is necessary to analyse the received data in detail. Based on the obtained data on the change

AU01	AU02	AU04	AU05	AU06	AU07	AU09	AU10	AU11	AU12	AU14	AU15	AU17	AU2	AU23	AU24	AU25	AU26	AU28	AU43
0.2015	0.439	0.401	0.61788225	0.10701495	0.0	0.151	0.304	0.4657	0.0783	0.25350195	0.26299292	0.49585113	0.0	0.2534673	0.5679667	0.6631179	0.232	0.1079816	0.0314216
0.2076	0.401	0.262	0.46185502	0.11505855	1.0	0.211	0.029	0.4652	0.0635	0.17666021	0.20985217	0.5158196	0.0	0.1839198	0.5350253	0.0356071	0.099	0.2477402	0.0238785
0.2466	0.420	0.203	0.44859222	0.15731426	0.0	0.154	0.036	0.4654	0.0816	0.28346816	0.11374501	0.49949518	0.0	0.2412251	0.5809660	0.0173820	0.211	0.1425058	0.0262596
0.2293	0.481	0.446	0.4106907	0.1294305	0.0	0.176	0.067	0.4691	0.0414	0.22750711	0.39544374	0.4942658	0.0	0.4365488	0.5564838	0.2244909	0.127	0.2749550	0.0307623
0.2444	0.496	0.221	0.4477415	0.17149642	1.0	0.236	0.213	0.4736	0.3165	0.42538813	0.1527553	0.5856195	0.0	0.4642386	0.5678026	0.0505657	0.120	0.1852337	0.0123650
0.2326	0.433	0.304	0.4697533	0.20796633	0.0	0.302	0.436	0.5015	0.2712	0.5791952	0.2966672	0.56989896	0.0	0.4474196	0.5566332	0.1065018	0.224	0.2329983	0.0253754
0.2003	0.377	0.394	0.42946061	0.25811613	0.0	0.211	0.229	0.4915	0.3773	0.40616524	0.2215647	0.63297725	0.0	0.2871313	0.5566332	0.9528224	0.193	0.1285169	0.0164890
0.2845	0.460	0.410	0.43317264	0.22879791	0.0	0.219	0.018	0.4717	0.3430	0.48396254	0.2967496	0.6152324	0.0	0.3518205	0.5715586	0.2203476	0.320	0.1303951	0.0176608
0.2020	0.392	0.440	0.41757047	0.06790614	0.0	0.168	0.079	0.4660	0.0235	0.30125552	0.77015877	0.51032454	0.0	0.6843622	0.4466270	0.2518663	0.104	0.2015777	0.0314326
0.1923	0.497	0.344	0.4596905	0.15082094	1.0	0.201	0.012	0.4625	0.1561	0.40891713	0.07942343	0.55068016	0.0	0.2907979	0.5576473	0.0686294	0.103	0.1606189	0.0216108
0.2605	0.543	0.252	0.44039127	0.1668334	0.0	0.157	0.036	0.4655	0.1147	0.43202522	0.09339548	0.5505572	0.0	0.2110465	0.5563051	0.7988006	0.151	0.1190917	0.0120105
0.2351	0.402	0.258	0.4399591	0.17693399	1.0	0.181	0.094	0.4703	0.2081	0.42365775	0.22079244	0.55712366	0.0	0.3096216	0.5723764	0.0650009	0.395	0.0487724	0.0145130
0.2408	0.461	0.186	0.40443555	0.21439436	0.0	0.165	0.080	0.4710	0.1510	0.40759522	0.12834156	0.56171167	0.0	0.3237983	0.5635962	0.5919624	0.268	0.0936644	0.0196618
0.3095	0.404	0.360	0.6381018	0.12647317	0.0	0.182	0.002	0.4672	0.1016	0.26860252	0.57722247	0.53266376	0.0	0.6143671	0.5754928	0.0219289	0.329	0.4192066	0.0255063
0.2561	0.438	0.416	0.6278245	0.14964405	0.0	0.156	0.213	0.4652	0.0631	0.35809502	0.50511247	0.5084298	0.0	0.7173407	0.441497	0.0172339	0.141	0.2304381	0.0168010
0.2988	0.495	0.322	0.36904383	0.21024275	0.0	0.167	0.612	0.4673	0.3006	0.48860523	0.3100148	0.5483622	0.0	0.6291554	0.5635962	0.7684691	0.414	0.4671391	0.0218129
0.2490	0.435	0.323	0.40085456	0.10973779	0.0	0.165	0.026	0.4623	0.0721	0.34405318	0.18081124	0.51806444	0.0	0.2826202	0.5665662	0.0405261	0.215	0.5945991	0.0721195
0.2232	0.383	0.452	0.44822925	0.099374086	1.0	0.176	0.018	0.4658	0.0658	0.12617053	0.21729803	0.6013818	0.0	0.3508865	0.5590743	0.1102696	0.137	0.1708438	0.0251759
0.2174	0.456	0.443	0.39579836	0.12659311	0.0	0.220	0.317	0.4662	0.0586	0.24809916	0.30578926	0.55373573	0.0	0.5047146	0.4970505	0.0033194	0.092	0.3262008	0.0426151
0.2166	0.450	0.310	0.44752178	0.13516106	0.0	0.200	0.011	0.4709	0.0745	0.20806004	0.3509115	0.5118353	0.0	0.1992240	0.5684011	0.1172482	0.079	0.2042432	0.0138056
0.2366	0.470	0.370	0.6993505	0.11621275	0.0	0.137	0.041	0.4642	0.0704	0.27254635	0.54328156	0.43498367	0.0	0.5185487	0.4470178	0.3267877	0.251	0.1667849	0.0235466
0.2282	0.456	0.457	0.64292365	0.13665833	1.0	0.129	0.182	0.4646	0.0836	0.28700185	0.43440574	0.4748872	0.0	0.5461804	0.5308797	0.5889336	0.313	0.1402956	0.0211922
0.2336	0.494	0.378	0.6648022	0.11174018	0.0	0.171	0.031	0.4631	0.0641	0.27300936	0.298793	0.43503386	0.0	0.3803767	0.5771656	0.8123944	0.193	0.3807165	0.0460994
0.2864	0.473	0.310	0.5822894	0.1354885	0.0	0.144	0.283	0.4678	0.0484	0.20962493	0.2360078	0.4630421	0.0	0.2857757	0.5511408	0.8279778	0.216	0.1601107	0.0506488
0.2073	0.482	0.287	0.42125228	0.14805456	0.0	0.208	0.108	0.4704	0.0941	0.4422492	0.41625413	0.49210352	0.0	0.3049377	0.5179967	0.5028346	0.170	0.1579078	0.0560562

Fig. 2. The result of processing a 25-s video

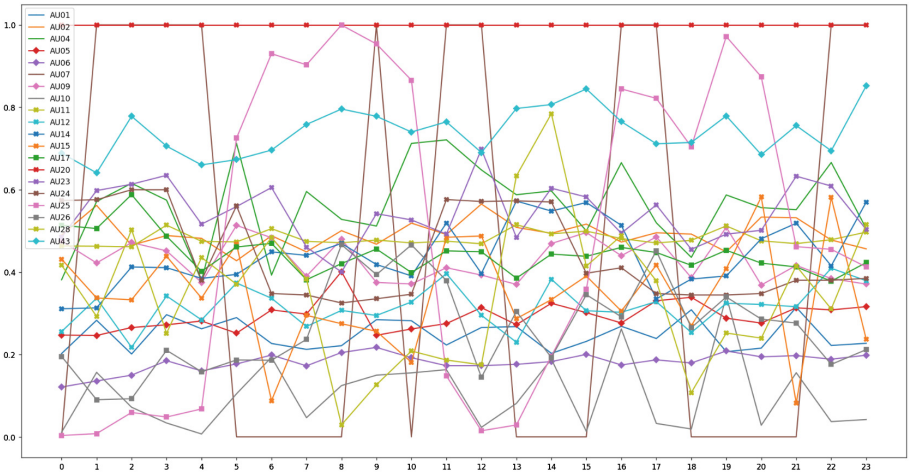

Fig. 3. Graph of facial muscle changes (AU) in a 24-s video clip

in AU (action units) provided by the analysis with the PyFeat Library, it is possible to plot the changes in facial facial muscles in the video fragment (Fig. 3).

For a more detailed definition of the manifestation of irony, graphs of changes in various muscle groups within the homonymous fragments with and without irony were compared (see Fig. 4).

We observed in many different video fragments the similar correlation between the presence of irony in a video fragment and the fact that AU 01 (raising the inner part of the eyebrows) and AU 02 (raising the outer part of the eyebrows) do not intersect with AU 04 (lowering the eyebrows). We also noticed that the facial expressions begin to change a little before the announcer was supposed to start pronouncing the ironic target

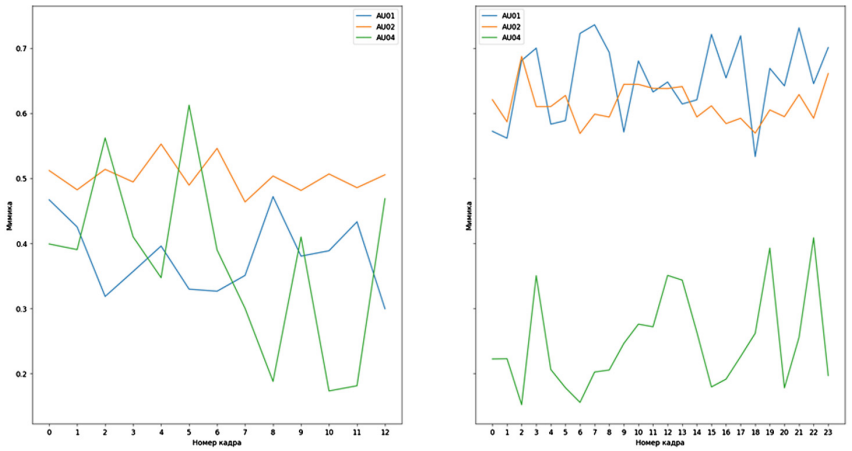

Fig. 4. Illustration of difference between eyebrow mimic movements in utterances without irony (left-hand example) and with irony (right-hand example); female speech; speaker ED.

fragment. Thus, we can conclude that the use of neural networks can reveal more patterns associated with the correlation of irony and changes in AU from different muscle groups.

Using Keras to Create a Neural Network. At the first stage of our work we defined a scheme for building a neural network was defined (Fig. 5):

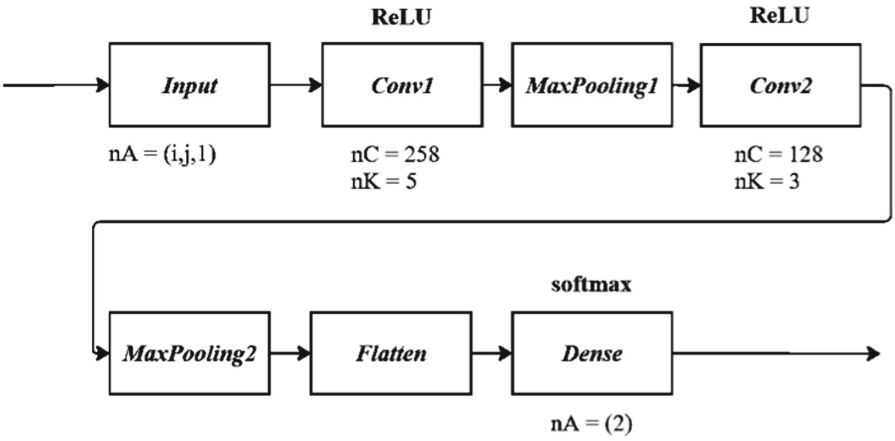

Fig. 5 Architecture of the neural network

Convolutional 2D layers serve as the implementation of convolutional layers, and a maximum-based pooling layer is located between them. When compiling the model, Adam is used as an optimizer, while the loss function is poisson or "poisson distribution". This loss function has experimentally proved to be the most effective. The summary method was also used, which allowed us to output general information about the created

network at the compilation stage, then the number of network parameters, the structure of the layers and the change in the output form were displayed.

Then we prepared the data for training and verification, and started the learning process. The data_path value or the "data path" for training was passed to the train_network method. Before training the network, all files along the specified path were processed and for each set of information about changes in facial muscles, the value of the presence or absence of irony was adjusted.

4.2 Results of Training Irony Recognizer

Before training, we divided the data set into two – training ("Train") and verification ("Test"). First, the network was trained on information from only 4 speakers – two female and two male, while all data about other speakers was placed in a test set. As a result of the training of the ANN during 20 epochs, the following data were obtained (Fig. 6).

```
Epoch 11/20
10/10 [==============================] - 0s 19ms/step - loss: 0.8024 - accuracy: 0.7104 - val_loss: 0.8304 - val_accuracy: 0.6275
Epoch 12/20
10/10 [==============================] - 0s 20ms/step - loss: 0.7886 - accuracy: 0.6869 - val_loss: 0.8362 - val_accuracy: 0.6013
Epoch 13/20
10/10 [==============================] - 0s 19ms/step - loss: 0.7803 - accuracy: 0.7407 - val_loss: 0.8362 - val_accuracy: 0.6275
Epoch 14/20
10/10 [==============================] - 0s 19ms/step - loss: 0.7703 - accuracy: 0.7071 - val_loss: 0.8433 - val_accuracy: 0.5621
Epoch 15/20
10/10 [==============================] - 0s 19ms/step - loss: 0.7643 - accuracy: 0.7677 - val_loss: 0.8449 - val_accuracy: 0.6078
Epoch 16/20
10/10 [==============================] - 0s 19ms/step - loss: 0.7492 - accuracy: 0.7744 - val_loss: 0.8512 - val_accuracy: 0.5686
Epoch 17/20
10/10 [==============================] - 0s 20ms/step - loss: 0.7374 - accuracy: 0.7710 - val_loss: 0.8614 - val_accuracy: 0.5098
Epoch 18/20
10/10 [==============================] - 0s 19ms/step - loss: 0.7308 - accuracy: 0.8316 - val_loss: 0.8483 - val_accuracy: 0.6209
Epoch 19/20
10/10 [==============================] - 0s 19ms/step - loss: 0.7316 - accuracy: 0.7845 - val_loss: 0.8809 - val_accuracy: 0.4967
Epoch 20/20
10/10 [==============================] - 0s 19ms/step - loss: 0.7119 - accuracy: 0.8653 - val_loss: 0.8617 - val_accuracy: 0.6405
```

Fig. 6. Data on network training during 20 epochs (epochs 11–20 are displayed)

The accuracy of the ANN on the training data reached 87%, on the verification data – 64% (Fig. 35). The very fact that training a neural network based on data containing changes in facial muscles, compared with the presence of irony in speech of only 4 speakers, is enough for the network to determine 64% correctly the presence of irony in speech of the other 5 speakers; this reveals the potential of deep learning in the studied field.

Then we changed the ratio of training data to verification data from 4 to 5 to 6 to 3, which increased the accuracy of the verification data, preventing the network from getting hung up on specific values (Fig. 7).

The shift in the data ratio had its effect. Despite the fact that the accuracy on the training data was only 80%, the accuracy on the verification data has already become 65%, while no changes have been made to the network architecture.

Thus, the results of the ANN work using convolutional layers showed that such an expressive feature of speech as irony can be associated with the mimic components of speech, namely changes in the position of AU (action units). The network, "knowing" how facial muscles behave when irony is manifested in four speakers, is already able to recognize it in 64% of cases for completely different people.

```
10/10 [==============================] - 1s 52ms/step - loss: 0.8384 - accuracy: 0.6229 - val_loss: 0.8397 - val_accuracy: 0.6471
Epoch 2/10
10/10 [==============================] - 0s 20ms/step - loss: 0.8362 - accuracy: 0.6465 - val_loss: 0.8269 - val_accuracy: 0.6471
Epoch 3/10
10/10 [==============================] - 0s 20ms/step - loss: 0.8260 - accuracy: 0.6229 - val_loss: 0.8263 - val_accuracy: 0.6471
Epoch 4/10
10/10 [==============================] - 0s 21ms/step - loss: 0.8215 - accuracy: 0.6229 - val_loss: 0.8260 - val_accuracy: 0.6471
Epoch 5/10
10/10 [==============================] - 0s 19ms/step - loss: 0.8200 - accuracy: 0.6229 - val_loss: 0.8279 - val_accuracy: 0.6471
Epoch 6/10
10/10 [==============================] - 0s 20ms/step - loss: 0.8074 - accuracy: 0.6229 - val_loss: 0.8253 - val_accuracy: 0.6471
Epoch 7/10
10/10 [==============================] - 0s 20ms/step - loss: 0.7987 - accuracy: 0.7071 - val_loss: 0.8272 - val_accuracy: 0.6405
Epoch 8/10
10/10 [==============================] - 0s 20ms/step - loss: 0.7916 - accuracy: 0.6633 - val_loss: 0.8335 - val_accuracy: 0.6144
Epoch 9/10
10/10 [==============================] - 0s 20ms/step - loss: 0.7763 - accuracy: 0.7239 - val_loss: 0.8325 - val_accuracy: 0.6536
Epoch 10/10
10/10 [==============================] - 0s 21ms/step - loss: 0.7640 - accuracy: 0.7980 - val_loss: 0.8408 - val_accuracy: 0.6536
```

Fig. 7. Data on network training during 10 epochs with a biased data ratio

5 Discussion and Conclusion

When expressing irony, a person uses various strategies for its implementation. It was observed both in audio and in video signals.

Acoustic characteristics can be used both individually and in combination. Some acoustic cues, such as phrase intensity, tempo, and vowel duration, could lead to false positives. For example, high intensity could be misinterpreted by the model as an indicator of irony, even though irony may be expressed in a calmer manner of speech if the person chose a different strategy for expressing it. In contrast, if irony was expressed through increasing the duration of the stressed vowel and slowing down the tempo without increasing intensity, the model might not recognize irony because it did not take these important criteria into account.

The complex nature of irony must be taken into account. In real life, its determination is not always possible with absolute certainty and accuracy. Thus, despite the presence of two classes, this problem is not a strictly binary classification. An important aspect is the degree of confidence with which one can say that an expression is ironic or not. Although features such as tempo, duration, and mel-cepstral coefficients are very important in determining irony, their interaction with other less relevant features (e.g., melodic range) can introduce noise into the model. As a result, the model could produce false positives based on features that do not provide significant information for detecting irony.

At the same time we can assume that the convolutional architecture is a successful way for building artificial neural networks to find a correlation between facial muscles changes of a person during speech with the manifestation of such an expressive feature as irony. Enlarging data will allow to precise the concrete action units that are the most important for the automatic irony detection in the video signal.

The future analysis will be focused on the complex and nuanced combinations of acoustic features with the co-speech facial expressions and their synchronization, as well as on modeling individual strategies of irony implementation basing on the acoustic and paralinguistic parameters of speaker's verbal behavior.

References

1. Ahsan, T., Jabid, T., Chong, U.P.: Facial expression recognition using local transitional pattern on Gabor filtered facial images. IETE Tech. Rev. **30**(1), 47–52 (2013). https://doi.org/10.4103/0256-4602.107339
2. Barbieri, F., Saggion, H., Ronzano, F.: Modelling sarcasm in Twitter, a novel approach. In: 2014, ACL, p. 50 (2014)
3. Bobe, A.S., Konyshev, D.V., Vorotnikov, S.A.: Sistema raspoznavaniya bazovyh emocij na osnove analiza dvigatel'nyh edinic lica [A system for recognizing basic emotions based on the analysis of facial motor units]. Inzhenernyj zhurnal: nauka i innovacii [Engineering Journal: Science and Innovation]. № 9. S. 7 (2016). https://doi.org/10.18698/2308-6033-2016-9-1530
4. Bryant, G., Fox Tree, J.: Is there an ironic tone of voice? In: Language and Speech, vol.48, pp. 257–277 (2008)
5. Candes, E., Demanet, L., Donoho, D., Ying, L.: Fast discrete curvelet transforms. Multiscale Model. Simul. **5**(3), 861–899 (2006). https://doi.org/10.1137/05064182X
6. Carvalho, P., Sarmento, L., Silva, M.J., De Oliveira, E.: Clues for detecting irony in user-generated contents: oh...!! it's so easy;-). In: Proceedings of the 1st International CIKM Workshop on Topic-Sentiment Analysis for Mass Opinion, pp. 53–56. ACM (2009).
7. Cheang, H., Pell, M.: The sound of sarcasm. Speech Commun. **50**(5), 366–381 (2008)
8. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. IEEE Trans. Patt. Anal. Mach. Intell. **23**(6), 681–685 (2001). https://doi.org/10.1109/34.927467
9. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models: their training and application. Comput. Vis. Image Understand. **61**(1), 38–59 (1995). https://doi.org/10.1006/cviu.1995.1004
10. Cutler, A.: On saying what you mean without meaning what you say. In: Proceedings from the 10th Regional Meeting of the Chicago Linguistic Society, pp. 117–123. CLS, Chicago (1974)
11. Davidov, D., Tsur, O., Rappoport, A.: Semi-supervised recognition of sarcastic sentencesin Twitter and Amazon. In: Proceedings of the Fourteenth Conference on Computational Natural Language Learning, pp. 107–116, Association for Computational Linguistics (2010)
12. Filik, R., Leuthold, H., Wallington, K., Page, J.: Testing theories of irony processing using eye-tracking and ERPS. J. Exp. Psychol. Learn. Memory Cogn. **40**(3), 811–828 (2014)
13. Gao Y., Leung, M.K.H.: Face recognition using line edge map. IEEE Trans. Patt. Anal. Mach. Intell. **24**(6), 764–779 (2002). https://doi.org/10.1109/TPAMI.2002.1008383
14. Gent, H., Adams, Ch., Tang., Y., Shih, Ch.: Deep learning for prosody-based irony classification in spontaneous speech. In: Interspeech 2022 Proceedings, pp. 3993–3997 (2022). https://doi.org/10.21437/Interspeech.2022-10978
15. Giora, R.: On irony and negation. Discourse Process. **19**(2), 239–264 (1995)
16. González-Ibánez, R., Muresan, S., Wacholder, N.: Identifying sarcasm in Twitter: a closer look. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers, vol. 2, pp. 581–586. Association for Computational Linguistics (2011)
17. Grishina, E.A.: Russkaya zhestikulyaciya s lingvisticheskoj tochki zreniya. Korpusnye issledovaniya [Russian gesticulation from a linguistic point of view. Corpus studies] — M., Publishing house YASK (2017)
18. Haverkate, H.: A speech act analysis of irony. J. Pragm. 14, 77–109 (1990)
19. Huang, Y.H., Huang, H.H., Chen, H.H.: Irony detection with attentive recurrent neural networks. In: Jose, J., et al. Advances in Information Retrieval. ECIR 2017. Lecture Notes in Computer Science(), vol. 10193. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-56608-5_45

20. Ivanko, S.L., Pexman, P.M.: Context incongruity and irony processing. Discourse Process. **35**(3), 241–279 (2003)
21. Joshi, A., Sharma, V., Bhattacharyya, P.: Harnessing context incongruity for sarcasm detection. In: Proceedings of 53rd Annual Meeting of the Association for Computational Linguistics, Beijing, China, p. 757 (2015)
22. Kochetkova, U., Skrelin, P., Evdokimova, V., Kachkovskaia, T.: The multimedia corpus of Russian ironic speech for phonetic analysis. In: Eismont, P., Khokhlova, M., Koryshev, M., Riekhakaynen, E. (eds.) Literature, Language and Computing. Springer, Singapore (2023). https://doi.org/10.1007/978-981-99-3604-5_19
23. Kochetkova, U., Skrelin, P., Evdokimova, V., Novoselova, D.: The speech corpus for studying phonetic properties of irony. In: Chernigovskaya, T., Eismont, P., Petrova, T. (eds.) Language, Music and Gesture: Informational Crossroads. Springer, Singapore (2021). https://doi.org/10.1007/978-981-16-3742-1_16
24. Li, S., Gong, D., Yuan, Y.: Face recognition using Weber local descriptors. Neurocomputing **122**, 272–283 (2013). https://doi.org/10.1016/j.neucom.2013.05.038
25. Martínez, A.M., Kak. A.C.: PCA versus LDA. IEEE Trans. Patt. Anal. Mach. Intell. **23**(2), 228–233 (2001). https://doi.org/10.1109/34.908974
26. Maynard, D., Greenwood, M.A.: Who cares about sarcastic tweets? Investigating the impact of sarcasm on sentiment analysis. In: Proceedings of LREC (2014)
27. Michael, St., Zahra, A.: Automatic sarcasm detection with textual and acoustic data. Int. J. Rec. Technol. Eng. **8**(4), 1357–1360 (2019)
28. Mishra, A., Bhattacharyya, P.: Predicting readers' sarcasm understandability by modeling Gaze behavior. In: Cognitively Inspired Natural Language Processing. Cognitive Intelligence and Robotics. Springer, Singapore (2018). https://doi.org/10.1007/978-981-13-1516-9_5
29. Negahdaripour, S.: Revised definition of optical flow: integration of radiometric and geometric cues for dynamic scene analysis. IEEE Trans. Patt. Anal. Mach. Intell. **20**(9), 961–979 (1998). https://doi.org/10.1109/34.713362
30. Niebuhr, O.: Rich reduction: sound-segment residuals and the encoding of communicative functions along the hypo-hyper scale. In: 7th Tutorial and Research Workshop on Experimental Linguistics, St. Petersburg, Russia, pp. 11–24 (2016)
31. Ekman, P., Friesen, W.: Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, Palo Alto (1978)
32. Schiller, D.: The face and the faceness: iconicity in the early facial semiotics of Paul Ekman, 1957–1978. Sign Systems Studies, [s. l.], **49**(3), 361–382 (2021). https://doi.org/10.12697/SSS.2021.49.3-4.06
33. Shan, C., Gong, S., McOwan, P.W.: Facial expression recognition based on local binary patterns: a comprehensive study. Image Vis. Comput. **27**(6), 803–816 (2009). https://doi.org/10.1016/j.imavis.2008.08.005
34. Skrelin, P., Kochetkova, U., Evdokimova, V., Novoselova, D.: Can we detect irony in speech using phonetic characteristics only? – looking for a methodology of analysis. In: Karpov, A., Potapova, R. (eds.) Speech and Computer. SPECOM 2020. Lecture Notes in Computer Science(), vol. 12335. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-60276-5_52
35. Zeng, Q., Li, A.-R.: A survey in automatic irony processing: linguistic, cognitive, and multi-X perspectives. In: Proceedings of the 29th International Conference on Computational Linguistics, pp. 824–836 (2022)