ISSN 2278-3091
Volume 9, No.1, January – February 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse91912020.pdf
https://doi.org/10.30534/ijatcse/2020/91912020

# Approbation of a method for studying the reflection of emotional state in children's speech and pilot psychophysiological experimental data

**Elena Lyakso[1], Nersisson Ruban[2], Olga Frolova [1], Viktor Gorodnyi [1], Yuri Matveev [3]**
[1]St. Petersburg University, Russia, lyakso@gmail.com;
[2]School of Electrical Engineering, VIT, Vellore, India, nruban@vit.ac.in
[3]ITMO University, St. Petersburg, Russia, matveev@mail.ifmo.ru

## ABSTRACT

The goal of the study is cross-linguistic research of 8-12 year old children's emotional speech classification by experts and automatically using the material of the Russian and Tamil languages. In the proposed project, this objective will be solved by a complex analysis of the emotional speech of informants with typical development in the age range of 8–12 years, using four emotional states “sadness, joy, anger and neutral”, to analyze what is relevant and new in connection with the lack of analogous data. Comparison of data on recognizing the emotional speech of children on the material of two languages belonging to different language groups will be made. Two approaches to emotional speech recognition are planned to use – by man and automatically, moreover, for automatic recognition, different analysis algorithms will be applied. The use of different algorithms for automated assessment of emotional states according to speech features that were previously used in analyzing a specific language is new and relevant. This will make it possible to determine their universality or specificity and, perhaps, based on them to develop an entirely new approach to analyzing emotional speech of children. The standardized approach to the collection of speech material will provide valid data on the recognition of child's emotions on the base of their speech characteristics.
In this paper, we describe the methodology of our study and preliminary results of a perceptual experiment.

**Key words:** child speech, cross-linguistic research, emotion classification, perception psychophysiological study, Russian and Tamil languages.

## 1. INTRODUCTION

The problem of reflecting a human’s state in the parameters of his voice and speech has been studied for a long time on the basis of different methodological approaches by specialists of various fields of knowledge – psychologists [1], physiologists [2-5], specialists in the field of speech technologies [6-9]. Conditionally, all studies on the emotional speech can be systematized in terms of the result for practical use: automatic recognition; emotional speech synthesis; application in psychological, physiological, and psychophysiological studies (traditionally, the results of these studies are theoretically significant); medical use for the rehabilitation of patients, which requires studying the speech characteristics in people in various physiological states and with different psychoneurological status and changing the characteristics of the speech message depending on the speaker’s emotional state; virtual training for people with the impaired emotional sphere. Lately – creating alternative communication systems; computer games and tutorials, human-computer interfaces.

Scenic speech, spontaneous and elicited emotional speech of adults, emotional speech of children has been researched. With regard to the latter, research is carried out mainly by psychologists and physiologists.

The possibility of reflecting emotions in voice, speech, and facial expressions has been studied for typically developing children [1, 10-15]. There is a methodological problem regarding the objectivity of evaluation in studies of emotional states. In this regard, a lot of works devoted to the emotion reflection in speech and vocalizations of children are based on the assessment of listeners’ opinions – adults listening to the vocalizations of infants and determining the states reflected in them [14, 16, 17]. Acoustic studies of emotional Russian speech using technical means to analyze the acoustic characteristics of sounds are linked to the names of Galunov V., Manerov V. [18, 19]. Studies on acoustics of emotional speech and vocalizations of children brought up in a Russian-speaking environment are single [3, 4, 5].

Automatic detection and classification of emotional child speech under natural conditions is a new area for research [3, 20]. Researchers focus on the use of methods for automatic recognition of child speech [e.g., 21-26].

The study of emotional speech aims to collect specialized speech corpora [27]. There are currently corpora of emotional and spontaneous child speech on the material of the Mexican Spanish language for children aged 7-13 years [28], British

This study is financially supported by the Russian Foundation for Basic Research (project 19-57-45008–IND_a) – for Russian researcher, and Department of Science and Technology (DST) (INTRUSRFBR382) - for Indian researcher.

English for children aged 4-14 years [29], the German language for 10-13-year-old children [29, 30]. Corpora containing the spontaneous speech of Italian children aged 8-12 years [31] and Swedish children aged 8-15 years [32] also include the emotional child speech. On the material of the Russian language, the first corpus of emotional spontaneous speech of children aged 3-7 years "EmoChildRu" [3] and database "AD_Child.Ru" for speech of atypical development children was created [33, 34]. These corpora contain video records of children's behavior, physiological data (hearing thresholds, phonemic hearing, etc.). The database "Adult-Child Speech Interaction" was created to study speech interaction in mother-child dyads depending on the neurological and psychophysiological state of the child [35]. It should be noted that the corpora of child speech in other languages (for example, for English, Swedish, German preschoolers [29] annotated for three states of emotional speech "comfort – neutral – discomfort" contain information only about the age of children (5-9 years); on the material of the French language [36] – about the gender of children.

Emotional speech contains both verbal and non-verbal information (tempo, rhythm, voice quality, etc.) more than calm, neutral speech. Non-verbal signals account for the rest of the useful information. Traditionally, studies of emotional speech use acoustic and perceptual analysis. Studies on voice biometrics note that the efficiency of expert perceptual identification depends on the quality of the speech material, the expert's qualification, his phonemic hearing [for example, 37].

The acoustic characteristics of speech are divided into two categories: phonetic and prosodic [38]. When determining the emotional state of a human by speech, prosodic characteristics are taken into account to a greater extent. To describe the prosodic characteristics of speech, it is informative to determine the shape of the intonation contour (IC) as a change of the pitch in time. Two approaches are generally used in IC assessing: visual approximation and instrumental analysis of pitch and its intensity. The pitch values correlate with the human's emotional state. When a man is sad, tired, or misses, the pitch values of the speech signal are low. The pitch range is maximum in a state of joy and reduced in a state of sadness. In a calm or neutral state, the average values of pitch range are even lower; the pitch range is rising in a state of anger. When a man is happy, angry, or scared, speech is loud, fast, with pronounced high-frequency components [39]. All of the listed characteristics of speech: tempo, pitch, energetic components, volume – are parameters used to classify emotions by speech [40].

In the speech analysis and its recognition, the acoustic features used in physiological studies are maximally related to the physiological and anatomical features of the human speech tract whereas works on automatic recognition use many derivatives sometimes not related to a real voice source. In the last decade, a large number of papers on the automatic recognition of human's emotional states by speech have been published. The specialized journal IEEE Transactions on Affective Computing is issued. This is an interdisciplinary and international journal whose purpose is to disseminate research results on the development of systems capable of recognizing, interpreting, and modeling human emotions. Annual competitions of the INTERSPEECH Computational Paralinguistics Challenge (ComParE) are held (http://www.compare.openaudio.eu/). Monographs were published [7, 41-44].

The most complete review of the current state of automatic emotion recognition is given in the paper by Rouast P.V. et al. (2019) [6]. The authors reviewed the literature from 2010 to 2017 with a focus on approaches (950 works on shallow and deep architectures) using deep learning methods to recognize the human's emotional states.

Like any pattern recognition system, the system for recognizing human emotional states by voice consists of modules of preprocessing a speech signal, extracting acoustic features, generating models of emotional states, and comparing these models (classification) with standards from a database formed during the training phase of the system.

In the INTERSPEECH Computational Paralinguistics Challenge (ComParE) series, the use of feature vectors of the OpenSMILE feature extraction toolkit is popular [45]. Analysis of a subset of 233 papers that use deep neural networks (DNN) found that they are used for training on spatial, temporal, and mixed features.

In most systems, feature vectors are extracted using the OpenSMILE feature extraction toolkit [45]. The Interspeech 2009 Emotion Challenge configuration file is supplied with OpenSMILE. One feature vector is extracted for each audio file. For example, for 16 low-level descriptors (LLD) and their deltas, as well as 12 derivatives, the resultant vector of 384 features is calculated.

Until the middle of the current decade, the traditional approach was the use of Gaussian Mixture Models (GMM), combinations of GMM-SVM, and i-vectors [46] for recognizing the wide-band acoustic events, including the speaker's emotional state recognition. Generative models, such as GMM, are often used to construct feature vectors in order to train classifiers, for example, SVM, to increase their efficiency. The paper [47] presents the results of studies on

the corpus of spontaneous emotional child speech FAU AIBO Emotion Corpus on improving emotion recognition on the base of GMM. In the paper [48], the cubic SVM classifier model is used for emotion identification from the Hindi language.

In addition to the classical GMM-SVM models and i-vectors, most researchers use WEKA toolkit for classification [49, 50], which implements the top-8 of classifiers by the efficiency: RBFNetwork, PART, MultiLayerPerceptron, SimpleLogistic, Logistic, NaiveBayes, ConjunctiveRule and Sequential Minimal Optimization (SMO).

The main idea is to replace GMM with DNN. DNN can be used to extract higher-level representations on the base of spatio-temporal spaces of handcrafted features. Reducing the dimension of handcrafted features using DNN can lead to improving accuracy on various databases, as happened in the case of i-vectors.

As a classifier in the case of spatial features, it is effective to use convolutional neural networks (CNN), into the input of which both spectrograms of speech signals and raw speech data can be fed. As a classifier in the case of temporal features, it is effective to use recurrent neural networks (RNN), into the input of which handcrafted features extracted on each frame are fed, which gives opportunity to simulate temporary changes at the low frame level. In general, the addition of RNN to simulate temporal changes leads to improving classification accuracy.

Another popular approach is to combine CNN to extract features and LSTM to extract a time context, including directly from a speech signal. For compatibility with the results of other papers, the Unweighted Average Recall (UAR) is used as a performance measure of the emotional state recognition [51].

The presented project is aimed at solving the main scientific problem of the effective comprehensive analysis of emotional child speech.

## 2. RELEVANCE

The study relevance is due to the theoretical significance of the problem of recognizing the emotional state according to voice and speech features. Solvable within the project, the problem of identifying speech message parameters, which are necessary and sufficient to determine the emotional state according to speech features, has scientific novelty and relevance determined by the theoretical and applied significance of the study. The theoretical significance of cross-linguistic research is to identify cultural peculiarities reflecting the emotional state in child speech features on the material of two different language families – Indo-European (Russian – Slavic group) and Dravidian languages (Tamil – Kannada group). The applied aspect is to obtain data on the reflection of different emotional states in voice and speech features on a sample of typically developing children and the possibility of further use of these data while working with children with developmental disabilities. Many developmental disorders or atypical development of children are accompanied with emotional disturbance which makes it difficult and, in some cases, impossible for the child to socially adapt to the society. The use of different algorithms for automatic assessment of emotional states on speech features, which were previously used in analyzing a specific language, is relevant. It will allow applying data on automatic recognition of the emotional states when creating interfaces and systems for teaching children with atypical development and developmental disorders accompanied with emotional disturbance.

## 3. METHOD

The proposed interdisciplinary cross-linguistic research will be implemented using classical and modern methods in the field of IT, linguistics, psychophysiology, which are logically combined to achieve a general goal.

The research will include 2 series of experiments: 1 – Psychophysiological series, 2 – Automatic classification of emotional speech of children.

### 3.1 Series 1 – Psychophysiological study

The goal of the study is to identify the parameters of a speech message which listeners (people) rely on while recognizing and classifying emotional child speech and to determine the acoustic features and linguistic peculiarities of child speech that are necessary and sufficient to define various emotional states of children according to their speech features on the material of the Russian and Tamil languages.

**Participants of the study:** Children aged 8-12 years will participate as test subjects.
1. Russian children for whom Russian is the native language – 100 children, 20 children in each age range (8, 9, 10, 11, 12 years old) – 10 boys and 10 girls each.
2. Indian children for whom Tamil is the native language – 100 children, 20 children in each age range (8, 9, 10, 11, 12 years old) – 10 boys and 10 girls each.
3. Groups of adults - participants of the perceptual experiment.

**Speech recording situations:** Emotional speech modelling situations in which children will be offered to say a standard set of phrases modelling of emotional states – sadness, joy,

anger, and a calm (neutral) emotional state. For each emotional state - 10 phrases. Russian children will pronounce similar phrases in Russian and English, Tamil children – in Tamil and English. Previously, the task will be explained to children, and the preliminary training will be conducted. Place of study: laboratory conditions, secondary schools.

Selecting speech material for children to pronounce: phrases carrying different emotional load will be chosen based on the active vocabulary of children, comics and cartoons. Children will have to do their best to say the given speech material – in a neutral way, simulating the state of sadness, joy, and anger as good as possible. In addition, spontaneous speech of children will be recorded.

**Speech analysis:**

1. Acoustic spectrographic analysis of the emotional speech of children, the features will be analyzed as follows:
Duration of utterances, words, vowels, stationary parts of vowels, pauses between words, phrases, utterances;
Pitch values and pitch range of utterances, words;
Pitch values and pitch range of formants and their amplitudes for stressed and unstressed vowels in words.

2. Linguistic analysis of spontaneous child speech, words frequency definition reflecting each emotional state.

3. Perceptual experiment. The participants of the perceptual experiment will be groups of adults – native Russian and Tamil speakers with different experience in interacting with children (professional experience, parents, household, no experience), of different age and the education level. For perceptual analysis, test sequences will be created. The tests will include the emotional speech of Russian children and emotional speech of Tamil children. Native speakers will be asked to recognize the emotional state, gender, age of the child while listening to test sequences. The method of creating test sequences and conducting the perceptual experiment has been designed in the framework of previous studies [3-5, 16, 51].

**Design of psychophysiological study:**

1. Selecting groups of test subjects with account of the child age and gender including the explanation of the goal and objectives of the study to parents and signing the informed consent to participate in the study by them. 2. Recording speech and behavior of children in different emotional states. 3. Preparing test sequences, selecting groups of adults listening to test sequences (listeners), conducting perceptual study. 4. Evaluating behavior of children by video recordings. 5. Perceptual study data processing and conducting spectrographic analysis of emotional child speech; text analysis. 6. Statistical data processing in order to establish the correlations between the emotional state and the speech features of children taking into account their age and gender. 7. Identifying similarities and differences in the emotional speech of children on the material of the Russian and Tamil languages.

### 3.2. Series 2 – Automatic recognition of emotional child speech

The method of automatic recognition of emotional child speech was developed and tested in our study [51]. The current research concentrates on an approach for Speech Emotion Recognition (SER) that is a branch of paralinguistics, related to speaker state and trait recognition (SSTR).

A general pipeline for SSTR is as follows. The speech signal is processed to extract informative features, which are fed to machine learning modules for acoustic- and/or language-based classification. In affective computing, arousal and valence are the two main dimensions along which continuous and dimensional affect is measured. Arousal is defined as physiological/psychological state of being (re-)active, while valence is the feeling of positiveness. The comfort classification can be thought as a three-state valence classification problem.

Acoustic models refer to affect classification models trained on features derived from acoustic/prosodic Low Level Descriptors (LLD). On the other hand, language-based classification employs linguistic information provided by automatic speech recognition (ASR). Emotion recognition performance can be improved by good ASR, by providing robust linguistic features to be fused with acoustic features. However, this is not always possible, since the recognition of affective (emotional) speech is itself very challenging. We use only acoustic models due to lack of Russian ASR trained on child speech, and since ASR trained on adult speech does not work on children's speech.

The state-of-the-art computational paralinguistics systems use large scale suprasegmental feature extraction via passing a set of summarizing statistical functionals (such as moments, extremes) over LLD contours. Pitch, Formants, Mel Frequency Cepstral Coefficients (MFCC), Modulation Spectrum, Relative Spectral Transform - Perceptual Linear Prediction (RASTA-PLP), Energy and variation features (i.e. Shimmer and Jitter) are frequently used as LLDs. In line with the state-of-the-art, we plan to extract acoustic features using the freely available openSMILE tool with a standard configuration file.

We extract openSMILE features with a configuration file used in the INTERSPEECH 2010 Computational Paralinguistics Challenge (ComParE) as baseline set. This feature set contains 1582 suprasegmental features obtained by passing 21 descriptive functionals (e.g. moments, percentiles,

regression coefficients) on 38 Low Level Descriptors (LLD) extracted from the speech signal. This configuration file is preferred over the one used in the 2015 edition of the ComParE Challenge, since in our recent work on a subset of this corpus [3], ComParE 2010 baseline set gave better results compared to the 2015 version, which is a 6373-dimensional acoustic feature set. As preprocessing, we apply z-normalization (i.e. standardization to zero-mean, unit variance) or min-max normalization to [0.1] range.

The most commonly employed classifiers in paralinguistics are Support Vector Machines (SVM), Artificial Neural Networks (ANN), Gaussian Mixture Models (GMM), and Hidden Markov Models (HMM). The state-of-the-art models of SER for the current databases are those trained with SVMs and Deep Neural Networks (DNN). From the ANN family, Extreme Learning Machines (ELM), which combine fast model learning with accurate prediction capability, are recently applied to multi-modal emotion recognition and computational paralinguistics, obtaining state-of-the-art results with modest computational resources. Consequently, we employ Kernel ELMs in this work, as well as a fast and robust classifier based on Partial Least Squares (PLS) regression. As a further baseline, we use SVMs.

We plan to report classification results in terms of accuracy and Unweighted Average Recall (UAR), which is introduced as performance measure in the INTERSPEECH 2009 Emotion Challenge. UAR is used to overcome the biased calculation of accuracy towards the majority class. It also gives a chance-level baseline performance as 1/K, where K is the number of classes. Therefore, in a 3-class problem, we have 33.3% chance-level UAR.

## 4. EXPERIMENTAL RESULTS

### 4.1 Psychophysiological study

The pilot perceptual study was carried out. The aim of the study was to reveal the possibility to recognize child's emotional state via speech by native Russian speakers and foreigners. The participants of the study were 10 Russian-speaking 8-12 year old children (4 boys and 6 girls were born and living in Saint Petersburg, Russia) and 142 adults as listeners (100 Russian-speaking medical students at the age of 19.1 ± 1.4 years, 40 men and 60 women and 42 foreign medical students at the age of 22 ± 3.9 years, 40 men and 2 women).

The place of audio and video recording of child's speech and behavior was the laboratory; the situation of the dialogue with the experimenter was used to obtain the child's spontaneous speech. The recordings were made by the "Marantz PMD660" recorder with external microphone "SENNHEIZER e835S" and video camera "SONY HDR-CX560E". Speech files were stored in Windows PCM format WAV, 44.100 Hz, 16 bits per sample; video files were in AVI format. The annotation of the child's speech material was made by 3 experts on three categories "comfort – neutral (calm) - discomfort".

Two test sequences included 90 child's speech signals (child's replies in dialogues with the experimenter) were created. The duration of pauses between the signals was 5 s, which allowed the listeners to fill in the forms with requested information. Test sequences were presented to listeners in an open field for 10-people groups. Listeners attributed the speech signal to "comfort - neutral - discomfort" categories.

All procedures were approved by the Health and Human Research Ethic Committee of Saint Petersburg State University and written informed consent was obtained from parents of the child participant. Russian-speaking listeners recognize the state of comfort (50%) in the speech of children better vs. the state of discomfort (32%) and neutral state (38%) (table 1).

**Table 1:** Confusion matrices for emotion recognition in the speech of children by Russian-speaking listeners

| State | Comfort | Neutral | Discomfort |
|---|---|---|---|
| **Comfort** | **50** | 34 | 16 |
| **Neutral** | 46 | **38** | 16 |
| **Discomfort** | 35 | 33 | **32** |

Foreign listeners recognize the state of comfort (43%) and neutral state (41%) worse than the state of discomfort (31%). Foreign students classify the state of Russian children via speech with lower probability comparing with Russian-speaking listeners (table 2).

**Table 2:** Confusion matrices for emotion recognition in the speech of children by foreign listeners

| State | Comfort | Neutral | Discomfort |
|---|---|---|---|
| **Comfort** | **43** | 37 | 20 |
| **Neutral** | 37 | **41** | 22 |
| **Discomfort** | 37 | 32 | **31** |

Russian-speaking women determine the state of comfort (54%, 43% - women and men, respectively) and discomfort (33% and 30%) better than men. Russian-speaking men recognize the neutral state better than women (43% and 36% - men and women, respectively) (table 3).

**Table 3:** Confusion matrices for emotion recognition in the speech of children by Russian-speaking men and women

| State | men | | | women | | |
|---|---|---|---|---|---|---|
| | **Comf** | **Neutr** | **Disc** | **Comf** | **Neutr** | **Disc** |
| **Comf** | **42** | 36 | 22 | **53** | 32 | 15 |
| **Neutr** | 38 | **42** | 20 | 48 | **37** | 15 |
| **Disc** | 30 | 40 | **30** | 37 | 30 | **33** |

### 4.2 Conclusions

The results of the pilot experiment showed the ability of adults to recognize the emotional state of children only on the basis of their speech. Women, who spend more time with children, better determine their emotional state than men. Foreign listeners are also capable of determining the emotional states of children while listening to their speech, but they are not guided by linguistic information, but by the characteristics of children's voices. However, the study showed that using only one situation, the child's dialogue with the experimenter to provoke the child's emotional state, is not enough. In further work, more different situations will be used in order to evoke emotional states in children and emotion manifestation in speech.

## 5. CONCLUSION

The theoretical significance of the study is to compare the cultural specificity of the emotional state manifestation in the speech features of children. A practical solution will be the possibility of using the data of the perceptual research on the adults' recognition of the emotional child's state for training specialists to better understand the states of the child and as normative data for working with children with developmental disorders of different etiologies. Our results can contribute to solving the interdisciplinary problem how to makes the learning process easier and more efficient [52].

Automatic recognition data of the emotional speech can be used for interfaces' creating for systems of teaching children with atypical development. Thus, according to the theoretical significance and practical application, the proposed project has no analogues and can be executed at a high professional level corresponding to the world level.

## REFERENCES


1. K. Izard. ***Human emotions***, N.Y.: Plenum Press, 1977, 496 p. https://doi.org/10.1007/978-1-4899-2209-0
2. E. Lyakso. **Characteristics of infant's vocalizations during the first year of life**, *International Journal of Psychophysiology*, Vol. 30, pp. 150-151, 1998.
3. E. Lyakso, O. Frolova, E. Dmitrieva, A. Grigorev, H. Kaya, A.A. Salah, and A. Karpov. **EmoChildRu: Emotional child Russian speech corpus**, *Lecture Notes in Computer Science*, Vol. 9319, pp. 144-152, Sep. 2015. https://doi.org/10.1007/978-3-319-23132-7_18
4. E. Lyakso and O. Frolova. **Emotion state manifestation in voice features: Chimpanzees, human infants, Children, Adults**, *Lecture Notes in Computer Science*, Vol. 9319, pp. 201-208, Sep. 2015. https://doi.org/10.1007/978-3-319-23132-7_25
5. O. Frolova and E. Lyakso. **Emotional speech of 3-years old children: Norm-risk-deprivation**, *Lecture Notes in Computer Science*, Vol. 9811, pp. 262-270, Aug. 2016. https://doi.org/10.1007/978-3-319-43958-7_4
6. P. V. Rouast, M. T. P. Adam, and R. Chiong. **Deep learning for human affect recognition: Insights and new developments**, *IEEE Transactions on Affective Computing*, Vol. 14, no. 8, pp. 1-20, Jan. 2019. https://doi.org/10.1109/TAFFC.2018.2890471
7. B. Schuller and A. Batliner. ***Computational Paralinguistics: Emotion, Affect and Personality in Speech and Language Processing***, Chichester, U.K.: John Wiley & Sons Ltd, 2013, 344 p. https://doi.org/10.1002/9781118706664
8. B. Schuller, S. Steidl, A. Batliner, A. Vinciarelli, K. Scherer, F. Ringeval, M. Chetouani, F. Weninger, F. Eyben, E. Marchi, M. Mortillaro, H. Salamin, A. Polychroniou, F. Valente, S. Kim, and M. Mortillaro. **The INTERSPEECH 2013 computational paralinguistics challenge: Social signals, conflict, emotion, autism**, in *Proc. Interspeech 2013, 14th Annual Conference of the International Speech Communication Association*, Lyon, France, 2013, pp. 148-152.
9. B. Schuller, S. Steidl, A. Batliner, E. Bergelson, J. Krajewski, C. Janott, A. Amatuni, M. Casillas, A. Seidl, M. Soderstrom, A.S. Warlaumont, G. Hidalgo, S. Schnieder, C. Heiser, W. Hohenhorst, M. Herzog, M. Schmitt, K. Qian, Y. Zhang, G. Trigeorgis, P. Tzirakis, and S. Zafeiriou. **The interspeech 2017 computational paralinguistics challenge: Addressee, cold & snoring. Computational Paralinguistics Challenge (ComParE)**, in *Proc. Interspeech 2017*, Stockholm, Sweden, 2017, pp. 3442-3446. https://doi.org/10.21437/Interspeech.2017-43
10. L. A. Camras and J. M. Shutter. **Emotional facial expressions in infancy**, *Emotion review*, Vol. 2, no. 2, pp. 120-129, Feb. 2010. https://doi.org/10.1177/1754073909352529
11. M. W. Sullivan and M. Lewis. **Contextual determinants of anger and other negative expressions in young infants**, *Developmental psychology*, Vol. 39, no. 4, pp. 693-705, July 2003. https://doi.org/10.1037/0012-1649.39.4.693
12. D. W. Leger, R. A. Thompson, J. A. Merritt, and J. J. Benz. **Adult perception of emotion intensity in human infant cries: Effects of infant age and cry**

**acoustics**, *Child Development*, Vol. 67, no. 6, pp. 3238-3249, Dec. 1996. https://doi.org/10.1111/j.1467-8624.1996.tb01911.x

13. G. E. Gustafson and J. A. Green. **On the importance of fundamental frequency and other acoustic features in cry perception and infant development**, *Child development*, Vol. 60, no. 4, pp. 772-780, Aug. 1989. https://doi.org/10.2307/1131017
14. J. Lindová, M. Špinka, and L. Nováková. **Decoding of baby calls: can adult humans identify the eliciting situation from emotional vocalizations of preverbal infants?** *PLoS One*, Vol. 10, no. 4, e0124317, Apr. 2015. https://doi.org/10.1371/journal.pone.0124317
15. E. Scheiner, K. Hammerschmidt, U. Jürgens, and P. Zwirner. **Acoustic analyses of developmental changes and emotional expression in the preverbal vocalizations of infants**, *Journal of Voice*, Vol. 16, no. 4, pp. 509-529, Dec. 2002. https://doi.org/10.1016/S0892-1997(02)00127-3
16. E. Lyakso, O. Frolova, A., and A. Grigorev. **A comparison of acoustic features of speech of typically developing children and children with autism spectrum disorders**, *Lecture Notes in Computer Science*, Vol. 9811, pp 43-50, Aug. 2016. https://doi.org/10.1007/978-3-319-43958-7_4
17. Y. Shimura and S. Imaizumi. **Adult attribution of the infant vocalizations**, *The Journal of the Acoustical Society of America*, Vol. 97, no. 5, p. 3240, Aug. 1995. https://doi.org/10.1121/1.411732
18. V. I. Galunov. **Some problems of the acoustic theory of speech production**, *Acoustical Physics*, Vol. 48, no. 6, pp. 749-751, Nov. 2002. https://doi.org/10.1134/1.1522045
19. V. Kh. Manerov. **Extralinguistic signals, human properties and states**, in *Proc. Speech, Emotions, and Personality. Symposium Materials and Reports [in Russian]*, pp. 21–31, 1978.
20. A. Batliner, S. Steidl, C. Hacker, and E. Nöth. **Private emotions versus social interaction: a data-driven approach towards analysing emotion in speech**, *User Modeling and User-Adapted Interaction*, Vol. 18, no. 1-2, pp. 175-206, Feb. 2008.
21. Y. Attabi and P. Dumouchel. **Emotion recognition from children's speech using anchor models**, in *Proc. Third Workshop on Child, Computer and Interaction, WOCCI 2012*, Portland, USA, 2012, pp. 82-86.
22. A. Potamianos, D. Giuliani, S. S. Narayanan, and K. Berkling. **Introduction to the special issue on speech and language processing of children's speech for child-machine interaction applications**, *ACM Transactions on Speech and Language Processing (TSLP)*, Vol. 7, no. 4, pp. 1-3, Aug. 2011. https://doi.org/10.1145/1998384.1998385
23. H. Meinedo and I. Trancoso. **Age and gender detection in the I-DASH project**, *ACM Transactions on Speech and Language Processing (TSLP)*, Vol. 7, no. 4, pp. 1-16, Aug. 2011. https://doi.org/10.1145/1998384.1998387
24. D. Bolaños, R. A. Cole, W. Ward, E. Borts, and E. Svirsky. **FLORA: Fluent oral reading assessment of children's speech**, *ACM Transactions on Speech and Language Processing (TSLP)*, Vol. 7, no. 4, pp. 1-19, Aug. 2011.https://doi.org/10.1145/1998384.1998390
25. S. Safavi, P. Jancovic, M. J. Russell, and M. J. Carey. **Identification of gender from children's speech by computers and humans**, in *Proc. Interspeech* 2013, Lyon, France, 2013, pp. 2440-2444.
26. S. Safavi, M. Russell, and P. Jančovič. **Identification of age-group from children's speech by computers and humans,** in *Proc. Fifteenth Annual Conference of the International Speech Communication Association*, Singapore, 2014, pp. 243-247.
27. D. Ververidis and C. Kotropoulos. **Emotional speech recognition: Resources, features, and methods**, *Speech communication*, Vol. 48, no. 9, pp. 1162-1181, Sept. 2006. https://doi.org/10.1016/j.specom.2006.04.003
28. H. Pérez-Espinosa, C. A. Reyes-García, and L. Villaseñor-Pineda. **EmoWisconsin: an emotional children speech database in Mexican Spanish**, in *Proc. International Conference on Affective Computing and Intelligent Interaction*, Berlin, Heidelberg, 2011, pp. 62-71. https://doi.org/10.1007/978-3-642-24571-8_7
29. A. Batliner, M. Blomberg, S. D'Arcy, D. Elenius, D. Giuliani, M. Gerosa, C. Hacker, M. Russell, S. Steidl, and M. Wong. **The PF_STAR children's speech corpus**, in *Proc. 9th European Conference on Speech Communication and Technology*, Lisbon, Portugal, 2005, pp. 2761-2764.
30. A. Batliner, S. Steidl, and E. Nöth. **Releasing a thoroughly annotated and processed spontaneous emotional database: the FAU Aibo Emotion Corpus**, in *Proc. of a Satellite Workshop of LREC 2008*, Marrakesh, 2008, pp. 28-31.
31. M. Gerosa, D. Giuliani, and F. Brugnara. **Acoustic variability and automatic recognition of children's speech**, *Speech Communication*, Vol. 49, no. 10-11, pp. 847-860, Oct.-Nov. 2007. https://doi.org/10.1016/j.specom.2007.01.002
32. L. Bell, J. Boye, J. Gustafson, M. Heldner, A. Lindström, and M. Wirén. **The Swedish NICE Corpus–Spoken dialogues between children and embodied characters in a computer game scenario**, in *Proc. Interspeech 2005 - Eurospeech, 9th European Conference on Speech Communication and Technology*, Lisbon, Portugal, 2005, pp. 2765-2768.
33. E. Lyakso, O. Frolova, and A. Karpov. **A New Method for Collection and Annotation of Speech Data of Atypically Developing Children**, *IEEE International Conference on Sensor Networks and Signal Processing (SNSP 2018)*, pp.175-180, Jan. 2019. https://doi.org/10.1109/SNSP.2018.00040

34. E. Lyakso, O. Frolova, A. Kaliyev, V. Gorodnyi, A. Grigorev, and Yu. Matveev. **AD-Child.Ru: Speech corpus for Russian children with atypical development**, *Lecture Notes in Computer Science*, Vol. 11658, pp. 299-308, July 2019. https://doi.org/10.1007/978-3-030-26061-3_31
35. E. Lyakso and O. Frolova. **Adult-child speech interaction: Speech database and psychophysiological experimental data**, *International Journal of Advanced Trends in Computer Science and Engineering*, Vol. 8, no. 5, pp. 2399-2407, Sept.-Oct. 2019. https://doi.org/10.30534/ijatcse/2019/81852019
36. A. Syssau and C. Monnier. **Children's emotional norms for 600 French words,** *Behavior Research Methods,* Vol. 41, no. 1, pp. 213-219, Feb. 2009. https://doi.org/10.3758/BRM.41.1.213
37. Y. Matveev. **The problem of voice template aging in speaker recognition systems**, *Lecture Notes in Computer Science,* Vol. 8113, pp. 345-353, Sept. 2013. https://doi.org/10.1007/978-3-319-01931-4_46
38. Z. T. Liu, M. Wu, W. H. Cao, J. W. Mao, J. P. Xu, and G. Z. Tan. **Speech emotion recognition based on feature selection and extreme learning machine decision tree**, *Neurocomputing*, Vol. 273, pp. 271-280, Jan. 2018. https://doi.org/10.1016/j.neucom.2017.07.050
39. S. Johar. **Psychology of voice**, *Emotion, affect and personality in speech,* Cham: Springer, 2016, pp. 9-15. https://doi.org/10.1007/978-3-319-28047-9_2
40. N. Ellouze. **Pitch and energy contribution in emotion and speaking styles recognition enhancement**, in *Proc. of the Multiconference on Computational Engineering in Systems Applications,* 2006, Vol. 1, pp. 97-100. https://doi.org/10.1109/CESA.2006.4281631
41. J. Swati. ***Emotion, Affect and Personality in Speech: The Bias of Language and Paralanguage***, NJ: Springer, 2016, 52 p. https://doi.org/10.1007/978-3-319-28047-9
42. A. Konar and A. Chakraborty. ***Emotion Recognition: A Pattern Analysis Approach***, Chichester, U.K.: John Wiley & Sons Ltd, 2015, 584 p. https://doi.org/10.1002/9781118910566
43. K. S. Rao and S. G. Koolagudi. ***Emotion Recognition using Speech Features***, N.Y.: Springer, 2013, 124 p. https://doi.org/10.1007/978-1-4614-5143-3
44. K. S. Rao, S. G. Koolagudi. ***Robust Emotion Recognition using Spectral and Prosodic Features***, N. Y.: Springer, 2013, 118 p. https://doi.org/10.1007/978-1-4614-6360-3
45. F. Eyben, M. Wollmer, and B. Schuller. **OpenSMILE: the Munich versatile and fast open-source audio feature extractor**, in *Proc. of the International Conference on Multimedia*, New York, United States, 2010, pp. 1459–1462. https://doi.org/10.1145/1873951.1874246
46. N. Dehak. **I-Vector representation based on GMM and DNN for audio classification,** *Keynote speech at Odyssey 2016, Speaker and Language Workshop*, Bibao, Spain, 2016.
47. Y. Attabi and P. Dumouchel. **Anchor models for emotion recognition from speech**, *IEEE Transactions on Affective Computing*, Vol. 4, no. 3, pp. 280-290, Aug. 2013. https://doi.org/10.1109/T-AFFC.2013.17
48. U. Jain, K. Nathani, N. Ruban, A. N. J. Raj, Z. Zhuang, and V. G. Mahesh. **Cubic SVM classifier based feature extraction and emotion detection from speech signals**, *IEEE International Conference on Sensor Networks and Signal Processing (SNSP 2018),* pp. 386-391. Jan. 2019. https://doi.org/10.1109/snsp.2018.00081
49. H. Witten and E. Frank. ***Data Mining: Practical Machine Learning Tools and Techniques***, ***Second Edition***, San Francisco: Elsevier, Morgan Kaufmann Publisher, 2005, 560 p.
50. S. Makki and F. Alqurashi. **An adaptive model for knowledge mining in databases "EMO_MINE" for tweets emotions classification**, *International Journal of Advanced Trends in Computer Science and Engineering*, Vol. 7, no. 3, pp. 52-60, May-June 2018. https://doi.org/10.30534/ijatcse/2018/04732018
51. H. Kaya, A. A. Salah, A. Karpov, O. Frolova, A. Grigorev, and E. Lyakso. **Emotion, age, and gender classification in children's speech by humans and machines**, *Computer Speech & Language,* Vol. 46, pp. 268-283, Nov. 2017. https://doi.org/10.1016/j.csl.2017.06.002
52. T. C. Sandanayake, A. M. Bandara. **Automated classroom lecture note generation using natural language processing and image processing techniques**, *International Journal of Advanced Trends in Computer Science and Engineering*, Vol. 8, no. 5, pp. 1920-1926, Sept.-Oct. 2019. https://doi.org/10.30534/ijatcse/2019/16852019