https://mirrors.creativecommons.org/presskit/buttons/80x15/png/by.png

http://dx.doi.org/10.35596/1729-7648-XXXX-XX-X-XX-XX

*Оригинальная статья  
Original paper*

УДК *004.934.2+534.784*

**СИСТЕМА РАСПОЗНАВАНИЯ ЭМОЦИОНАЛЬНОГО СОСТОЯНИЯ ДИКТОРА ПО РЕЧИ**

КРАСНОПРОШИН Д.В. ВАШКЕВИЧ М.И.

*Белорусский государственный университет информатики и радиоэлектроники*

*(г. Минск, Республика Беларусь)*

*Поступила в редакцию*

© Белорусский государственный университет информатики и радиоэлектроники, 2021

**Аннотация.** В работе описан подход к построению системы….

**Ключевые слова:** голосовой сигнал, МЧКК, БЧКК, патология голоса

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

**Для цитирования.** Вашкевич М.И., Лихачев Д.С., Азаров И.С. Система анализа и классификации голосового сигнала на основе пертурбационных параметров и кепстрального представления в психоакустических шкалах. Доклады БГУИР. 2021; \*\*(\*): \*\*\*-\*\*\*.

**VOICE ANALYSIS AND CLASSIFICATION SYSTEM BASED ON PERTURBATION PARAMETERS AND CEPSTRAL PRESENTATION IN PSYCHOACOUSTIC SCALES**

MAXIM.I. VASHKEVICH, DENIS S. LIKHACHOV, ELIAS S. AZAROV

*Belorussian state university of informatics and radioelectronics*

*P.Brovki str., 6, Minsk, 220013, Republic of Belarus*

*Submitted*

© Belarusian State University of Informatics and Radioelectronics, 2021

**Abstract.** The paper describes an approach to design a system for analyzing and classification of a voice signal based on perturbation parameters and cepstral presentation. Two variants of the cepstral presentation of the voice signal are considered: based on mel-frequency cepstral coefficients (MFCC) and based on bark-frequency cepstral coefficients (BFCC). The work used a generally accepted approach to calculating the MFCC based on time-frequency analysis by the method of discrete Fourier transform (DFT) with summation of energy in subbands. This method approximates the frequency resolution of human hearing, but has a fixed temporal resolution. As an alternative, a variant of the cepstral presentation based on the BFCC has been proposed. When calculating the BFCC, an warped DFT-modulated filter bank was used, which approximates the frequency and temporal resolution of hearing. The aim of the work was to compare the effectiveness of the use of features based on the MFCC and BFCC for the designing systems for the analysis and classification of the voice signal. The results of the experiment showed that in the case of using acoustic features based on the MFCC, it is possible to obtain a voice classification system with an average recall of 80.6%, and in the case of using features based on the BFCC, this metric is 83.7%. With the addition of the set of MFCC-features with perturbation parameters of the voice, the average recall of the classification increased to 94.1%, with a similar addition to the set of BFCC-features, the average recall of the classification increased to 96.7%.

**Keywords:**voice signal, MFCC, BFCC, vocal pathology

**Conflict of interests.** The authors declare no conflict of interests.

**For citation.** Vashkevich M.I., Likhachov D.S., Azarau I.S. Voice analysis and classification system based on perturbation parameters and cepstral presentation in psychoacoustic scales. Doklady BGUIR. 2021; \*\*(\*): \*\*\*-\*\*\*.

**Введение**

1) Обозначить актуальность

2) Области применения

Распознавание эмоций по речи используются в здравоохранении. Распознавание негативных эмоций таких как стресс, злость, усталость является важным аспектом сточки зрения обеспечения безопасности дорожного движения с применением интеллектуальных транспортных средств, поскольку позволяет им реагировать на эмоциональное состояние водителя.

3) Существующие ограничения

4) Цели и задачи исследования

**Пертурбационные параметры голоса**

Пертурбационные параметры также рассчитываются исходя контура частоты основного тона (ЧОТ). К этой группе относят: 1) частотный диапазон фонации (англ. *PFR – phonatory frequency range*); 2) среднеквадратичное отклонение ЧОТ – ; 3) энтропия периодов ОТ (англ. *PPE – pitch period entropy*) [4]; 4) индекс патологичности вибрато (англ. *PVI – pathology vibrato index*). Используемые в работе пертурбационные параметры приведены в таблице 1. Более подробное их описание можно найти в работе [5].

**Таблица 1.** Пертурбационные параметры голоса

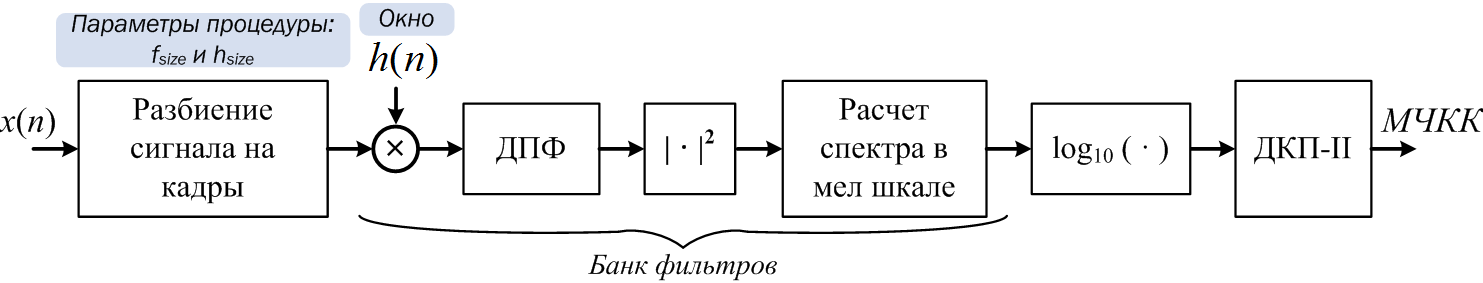
**Table 1.** Perturbation voice parameters

|  |  |  |
| --- | --- | --- |
| Группа параметров | Число параметров | Названия параметров |
| Частотная пертурбация | 5 | , , , , |
| Амплитудная пертурбация | 5 | , , , , |
| Пертурбация контура ЧОТ | 4 | , , |
| **Всего** | 14 |  |

**Кепстральное представление голоса в психоакустических шкалах**

В данном разделе рассматривается кепстральное представление голосового сигнала, получаемое на основе спектрального анализа сигнала в психоакустически мотивированной частотной шкале. Анализируется широко применяемое для описания голосового сигнала мел-частотное кепстральное представление [6], которое сравнивается с предлагаемым в работе барк-частотным кепстральным представлением, получаемым на основе неравнополосного ДПФ-модулированного банка фильтров.

Расчет мел-частотных кепстральных коэффициентов (МЧКК) относится к методам кратковременного анализа голосового сигнала, которые предполагают разбиение сигнала на кадры анализа. Как правило, в интервале от 10 до 30 мс голосовой сигнал можно считать стационарным. Схема вычисления МЧКК показана на рис. 1.



**Рис. 1.** Схема вычисления мел-частотных кепстральных коэффициентов (МЧКК)

**Fig. 1.** Scheme for calculating mel-frequency cepstral coefficients (MFCC)

**Речевая база**

1) Описание RAVDESS

При проведении исследования в качестве исходного набора данных использовался Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) [9]. RAVDESS содержит 7356 записей 24 актеров (12 мужчин, 12 женщин). Все актеры произвели 104 различных вокализации, состоящих из 60 устных высказываний и 44 песенных высказывания. Каждая из 104 вокализаций была экспортирована для создания трех отдельных модальных звуковых условиях: аудио-видео (лицо и голос), только видео (лицо, но без голоса) и только аудио (голос, но без лица). На каждого актера приходилось 312 файлов (104 × 3). Записи одного участника были потеряны по техническим причинам (132 файла). Таким образом, 24 × 312–132 = 7356 файлов. Этот набор состоит из 4320 записей речи и 3036 песен. Актеры озвучили две разных фразы (в речи и песни). Две фразы произносились с восемью эмоциональными окрасками (нейтральность, спокойствие, счастье, грусть, злость, страх, удивление и отвращение). В случае с песнями использовалось шесть эмоциональных окрасок (нейтральность, спокойствие, счастье, грусть, злость и страх). Все эмоциональные состояния, кроме нейтрального, озвучивались на двух уровнях эмоциональной громкости (нормальная и повышенная). Актеры повторяли каждую вокализацию дважды.

В рамках данной работы будет использована только часть датасета RAVDESS, а именно RAVDESS Emotional speech audio. Эта часть RAVDESS содержит 1440 файлов в формате wav (16 бит, 48 кГц): 60 записей на каждого из 24-х профессиональных актера (12 мужчин, 12 женщин). Фразы с нейтральным североамериканским акцентом. Речевые эмоции включают выражения нейтральности, спокойствия, счастья, грусти, гнева, страха, удивления и отвращения. Все эмоциональные состояния, кроме нейтрального, озвучивались на двух уровнях эмоциональной громкости (нормальная и повышенная). Актеры повторяли каждую вокализацию дважды.

Для построения системы по распознаванию эмоций в речи требуется провести предобработку исходных данных. Основной задачей предобработки является удаление шума, повышение высоких частот сигнала и получение плоского частотного спектра сигналов, а также частотных характеристик.

Как было упомянуто ранее, среди проблем связанных с обработкой речи особое место занимает выделение и выбор признаков. Различные аудио признаки позволяют описывать различные аспекты звукового сигнала для решения разного рода прикладных задач.

2) Подход к описанию эксперимента (evaluation design)

*We consider it necessary to establish a common evaluation design that we introduce and explain along with the paper that consists of a subject-wise 5-CV strategy using the eight emotions recorded in the RAVDESS dataset. взято из [1\*\*] нужно посмотреть их обоснование и применить их стратегию оценивания результата.*

**Результаты и их обсуждение**

Эксперимент проводился в три этапа: 1) подготовка обучающей выборки; 2) ранжирование признаков; 3) обучение и тестирование классификатора с использованием различного числа признаков.

**Таблица 3.** Максимальная средняя полнота классификации при ограничении числа признаков .

**Table 3.** The maximum average recall of the classification achieved when the number of features  is limited.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Набор признаков | Средняя полнота, % (В скобках указано число признаков, при котором достигается средняя полнота) | | | | | |
|  |  |  |  |  |  |
| MFCC-10 | 76,9 (5) | 77,2 (15) | – | – | – | – |
| MFCC-20 | 61,9 (9) | 62,4 (17) | 67,6 (30) | 70,1 (40) | 74,0 (48) | – |
| MFCC-40 | 78,0 (9) | 80,6 (12) | – | – | – | – |
| BFCC-10 | 72,2 (10) | 73,5 (11) | – | – | – | – |
| BFCC-20 | 77,1 (10) | 77,4 (13) | 80,0 (28) | – | – | – |
| BFCC-40 | **83,7** (6) | – | – | – | – | – |

Полученные в эксперименте результаты суммируются в табл. 3, где приведена оценка средней полноты классификации в зависимости от набора кепстральных признаков, при ограничении на число признаков. Такое ограничение обусловлено тем, что для классификации всегда желательно иметь меньшее число признаков, но обеспечивающих лучшее качество. Лучший результат достигается при использовании набора признаков BFCC-40 (см. табл. 3).

**Заключение**

В работе предложен метод вычисления барк-частотных кепстральных коэффициентов (БЧКК), основанный на использовании неравноплосного ДПФ-модулированного банка фильтров, аппроксимирующего частотно-временное разрешение слуха человека. Произведено сравнение предложенных БЧКК с широко распространенными мел-частотными кепстральными коэффициентами (МЧКК) в отношении эффективности построения на их основе системы анализа и классификации голосовых сигналов. Проведенные эксперименты по построению системы классификации голосов пациентов с неврологическим заболеванием БАС показали эффективность применения надсегментных БЧКК признаков. Среди классификаторов, использующих набор кепстральных признаков лучший результат (средняя полнота 83,7%) достигнут LDA-классификатором, использующим 6 надсегментных БЧКК признаков, отобранных методом LASSO. Среди классификаторов, использующих набор кепстральных признаков, объединенных с пертурбационными параметрами голоса, лучший результат (средняя полнота 96,7%) достигнут LDA-классификатором, использующим 45 надсегментных БЧКК признаков, отобранных методом LASSO.

**Список литературы / References**

1. Luna-Jiménez, Cristina, et al. "Multimodal emotion recognition on ravdess dataset using transfer learning." Sensors 21.22 (2021): 7665.
2. Likhachov, D., Vashkevich, M., Azarov, E., Malhina, K., Rushkevich, Y. “A mobile application for detection of amyotrophic lateral sclerosis via voice analysis”. *International Conference on Speech and Computer*, 2021, Springer, Cham, pp. 372-383.
3. Benba A., Jilbab A., Hammouch A., “Discriminating between patients with Parkinson’s and neurological diseases using cepstral analysis”, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 2016; 24 (10): 1100–1108.
4. Tsanas A., et al. "Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease." *IEEE Transactions on Biomedical Engineering* 59.5 (2012): 1264-1271.
5. Vashkevich M., Rushkevich Y. “Classification of ALS patients based on acoustic analysis of sustained vowel phonations,” *Biomedical Signal Processing and Control*, 2021; 65: 1-14.
6. Huang X., Acero A., Hon H.­W. *Spoken language processing: A guide to theory, algorithm, and system development*. Prentice hall PTR, 2001, p. 980.
7. Bielawski K., Petrovsky A., Proposition of minimum bands multirate noise reduction system which exploits properties of the human auditory system and all-pass transformed filter bank. *IEEE Workshop Signal Processing,* 2001, pp. 65-70.
8. James G. et al. *An introduction to statistical learning with applications in R*. NewYork: Springer, 2013. p. 426.
9. Vashkevich M., Petrovsky A. Rushkevich Y. Bulbar ALS detection based on analysis of voice perturbation and vibrato. IEEE *International Conference on* *Signal Processing: Algorithms, Architectures, Arrangements, and Applications*, 2019, pp. 267-272.

**Вклад авторов**

Вашкевич М.И. цель и задачи исследования, предложил идею барк-частотного кепстрального представления голосового сигнала, выполнил программную реализацию расчета БЧКК, принимал участие в подготовке текста статьи и интерпретации результатов экспериментов. Лихачев Д.С. выполнил программную реализацию расчета МЧКК, участвовал в подготовке программной базы для эксперимента. Азаров И.С. предложил идею совместного использования кепстральных признаков и пертурбационных параметров, принимал участие в подготовке текста статьи и интерпретации результатов экспериментов.

**Authors contribution**

Vashkevich M.I. determined the purpose and objectives of the study, proposed the idea of the bark-frequency cepstral representation of the voice signal, carried out the software implementation of the BFCC calculation, took part in the preparation of the text of the article and the interpretation of the experimental results. Likhachov D.S. carried out the software implementation of the calculation of the MFCC, participated in the preparation of the software tools for the experiment. Azarov I.S. proposed the idea of the joint use of cepstral features with perturbation parameters, took part in the preparation of the text of the article and interpretation of the experimental results.

|  |  |  |
| --- | --- | --- |
| **Сведения об авторах**  Вашкевич М.И., к.т.н., доцент кафедры электронных вычислительных средств (ЭВС) Белорусского государственного университета информатики и радиоэлектроники (БГУИР).  Лихачев Д.С., к.т.н., доцент кафедры ЭВС БГУИР  Азаров И.С., д.т.н., заведующий кафедрой ЭВС БГУИР |  | **Information about the authors**  Vashkevich M.I., PhD, Associate Professor of Computer Engineering Department of Belarusian State University of Informatics and Radioelectronics (BSUIR)  Likhachov D.S., PhD, Associate Professor of Computer Engineering Department of BSUIR  Azarov E.S. D.Sci., Head of Computer Engineering Department of BSUIR. |
| **Адрес для корреспонденции**  220013, Республика Беларусь, г. Минск, ул. П. Бровки, д. 6, Белорусский государственный университет информатики и радиоэлектроники  тел. +375-17-293-84-78;  e-mail: sanko@bsuir.by  Вашкевич Максим Иосифович |  | **Address for correspondence**  220013, Republic of Belarus, Minsk, P. Brovki str., 6, Belarusian State University of Informatics and Radioelectronics  tel. +375-17-293-84-78;  e-mail: vashkevich@bsuir.by  Vashkevich Maksim Iosifovich |