

СТАТЬИ

ВЫРАЖЕНИЕ ЭМОЦИЙ В РАЗНЫХ ЯЗЫКАХ: ВОЗМОЖНОСТИ МЕЖЪЯЗЫКОВОГО ПЕРЕНОСА МОДЕЛЕЙ РАСПОЗНАВАНИЯ

В.И. Лемаев, Н.В. Лукашевич

*Московский государственный университет имени М.В. Ломоносова, Москва,
Россия; vladzhkv98@mail.ru, louk_nat@mail.ru*

Аннотация: Задача автоматического распознавания эмоций в разговорной речи в последнее время приобретает все больше популярности, в первую очередь благодаря появлению больших датасетов размеченных данных и развитию моделей машинного обучения. Наиболее распространенным подходом для решения данной задачи на текущий момент является применение речевых моделей с архитектурой трансформера, при этом классификация обычно производится на основе небольшого набора из 4–9 базовых эмоций. Однако даже проявление этих базовых эмоций у разных людей может в значительной степени отличаться в зависимости от языка говорящего и его культурного окружения.

В настоящей работе исследуются возможности современных моделей машинного обучения, обученных на речевых эмоциональных данных разных языков, в частности то, какие лингвистические характеристики языков наибольшим образом влияют на их поведение. Для этого было проведено сравнение между собой качества автоматической классификации эмоций при применении речевых моделей, обученных на данных разных языков с дальнейшим их переносом на русский язык. В исследовании использовалась речевая модель HuBERT, отдельно обученная на данных польского, китайского и японского языков и затем протестированная на данных русского языка. В качестве обучающих данных для иностранных языков были соответственно взяты датасеты эмоциональной речи nEMO, ESD и JVNv. Тестирование обученных моделей производилось на данных русского датасета Dusha. Полученные результаты указывают на первоочередную зависимость речевой модели от просодических характеристик данных, в то время как различия в синтаксисе языков не оказывают большого влияния на качество классификации.

Ключевые слова: распознавание эмоций, распознавание речи, обработка естественного языка, машинное обучение, трансформеры, межъязыковой перенос моделей, HuBERT

doi: 10.55959/MSU0130-0075-9-2025-48-04-1

Для цитирования: Лемаев В.И., Лукашевич Н.В. Выражение эмоций в разных языках: возможности межъязыкового переноса моделей распознавания // Вестн. Моск. ун-та. Серия 9. Филология. 2025. № 4. С. 9–20.

EMOTIONAL EXPRESSION IN DIFFERENT LANGUAGES: CROSS-LINGUAL MODEL TRANSFER

V.I. Lemaev, N.V. Loukachevitch

*Lomonosov Moscow State University, Moscow, Russia; vladzhkv98@mail.ru,
louk_nat@mail.ru*

Abstract: The task of Speech Emotion Recognition has recently gained in popularity, mainly due to the increase of the available data and the development of the new machine learning algorithms. Currently, the most common approach to solving this task is to use transformer-based speech models, with the goal to determine whether the phrase belongs to one of 4–9 basic emotions. However, the way people express even these basic emotions can differ based on their native language and cultural background.

This paper analyses the capabilities of modern machine learning models, trained on emotional speech data from different languages, with the main goal to examine the effect of language's innate linguistic features on a model's behavior. For that, we compared the results of a few models, trained on the emotional data of different languages and transferred to the Russian language. We trained the HuBERT speech model on the task of speech emotion recognition, using training data from nEMO, ESD and JVNВ emotional speech datasets, which correspond to Polish, Chinese and Japanese languages respectively. The trained models then were tested on Russian emotional speech dataset Dusha. The obtained results show prosodic features being the main contributor to the model's performance, while differences in syntax generally not affecting it.

Keywords: emotion recognition; speech recognition; natural language processing; machine learning; transformers; cross-lingual transfer; HuBERT

For citation: Lemaev V.I., Loukachevitch N.V. (2025) Emotional Expression in Different Languages: Cross-Lingual Model Transfer. *Lomonosov Philology Journal. Series 9. Philology*, no. 4, pp. 9–20.

Введение

Язык и эмоции — два фундаментальных аспекта человеческого опыта, тесно переплетенные между собой. Язык служит основным средством выражения мыслей и чувств, а эмоции, в свою очередь, влияют на то, как человек воспринимает и использует язык. При этом как между собой различаются сами языки, так и проявление

эмоций в них может отличаться из-за культурных особенностей, способов выражения чувств, а также наличия или отсутствия в языке определенных слов для эмоций [Lindquist 2021].

Попытки автоматического распознавания человеческих эмоций на основе речи предпринимаются уже довольно долгое время. Понимание внутреннего эмоционального состояния человека может оказаться существенно полезным в областях взаимодействия человека и компьютера, в областях психологии и маркетинга, в области здравоохранения, например для постановки диагноза [Hansen et al. 2022], а также в области обеспечения общественной безопасности [Lefter, Jonker 2017]. Наиболее распространенным подходом к распознаванию эмоций в речи на данный момент является обучение моделей машинного обучения со специальной архитектурой — *трансформеров* [Vaswani et al. 2017], — которые анализируют аудиосигнал и/или текстовую транскрипцию речи, чтобы определить эмоциональное состояние говорящего. Сам же процесс обучения выполняется на большом количестве заранее собранных данных, причем метод их сбора часто отличается в зависимости от языка и самих исследователей.

Для русского языка имеются работы, указывающие на приоритет просодических и акустических характеристик речи при решении задачи автоматического распознавания эмоций [Верхоляк 2021], а также работы, посвященные выбору оптимального набора базовых эмоций для обучения языковых моделей [Sboev et al. 2021].

Основная цель данной статьи состоит в исследовании возможностей применения моделей, обученных распознавать эмоции в речи на данных одного языка, к данным другого языка с минимальным дообучением. В частности, наиболее интересным моментом является то, как ведут себя модели разных языков по сравнению друг с другом, проявляются ли у них те же лексические, грамматические и просодические особенности, свойственные носителям языка. В качестве речевой модели для исследования использовалась модель HuBERT [Hsu et al. 2021], обученная на данных иностранных языков, после чего она была протестирована на данных русского языка, в роли которых выступил датасет эмоциональной речи Dusha [Konratenko et al. 2022]. В качестве иностранных языков для обучения модели были выбраны английский, польский, китайский и японский языки из-за наличия для них качественно собранных датасетов эмоциональной речи. Исследование придерживается подхода zero-shot, тем самым при обучении модели не были использованы данные целевого (русского) языка.

Язык и эмоции

Выражение эмоций через язык позволяет людям лучше понимать друг друга и может осуществляться сразу на нескольких уровнях:

- *Лексический*: говорящий напрямую обозначает испытываемую им эмоцию с помощью слов, например «Мне грустно»;
- *Грамматический*: эмоция может выражаться с помощью определенных грамматических конструкций, например изменения порядка слов в русском языке: ср. предложение «Ты мне не веришь?», выражающее скорее нейтральный вопрос, и предложение «Не веришь ты мне?», выражающее удивление и возмущение;
- *Просодический*: на эмоциональность высказывания может влиять акцентирование и интонация, при этом одна и та же фраза в зависимости от них может передавать разные эмоции. Например, «Ну да...» может означать согласие, сарказм или разочарование.

Выражение эмоций разнообразно и сильно варьируется между языками. Наиболее заметные отличия часто проявляются в лексике, когда в одном языке для описания определенной эмоции присутствует слово, не имеющее аналогов в других языках, например, португальское понятие «саудаде», обозначающее ностальгию или тоску по чему-то безвозвратно утраченному.

Отдельная проблема возникает, когда необходимо получить упорядоченную классификацию всех эмоций, которые может испытывать человек. Типичные наборы эмоций, используемые в теоретических и практических областях психологии, насчитывают около 300 разнообразных эмоций и их вариаций. Однако в это число также входит множество вариаций близких по значению эмоций и их производных. Когда говорят о базовых эмоциях, универсальных для большинства языков, исследователи обычно выделяют 6 *архетипических* эмоций, на которые часто принято раскладывать все остальные эмоции — это гнев, страх, удивление, радость, отвращение и грусть [Cowie et al. 2001]. Определение такого небольшого набора необходимо в том числе и для задач автоматического распознавания эмоций. Но даже внутри этих базовых эмоций в разных языках наблюдаются небольшие флуктуации значения [Jackson et al. 2019].

Помимо номинативной классификации эмоций, многие исследователи также классифицируют их на основе двумерного пространства. Первой осью такого пространства выступает понятие «валентности» (англ. valence), которое описывает выражаемую эмоцию в терминах положительности/отрицательности. За вторую же ось отвечает понятие «интенсивность» (англ. arousal), которое описывает степень проявления выражаемой эмоции. Такое двумер-

ное пространство в высокой степени коррелирует с названными выше архетипическими эмоциями. Так, возбуждение симпатической нервной системы происходит при выражении эмоций высокой интенсивности — гнева, страха и радости — и приводит к повышенному сердечному ритму, напряжению мышц, повышенному кровяному давлению и сухости во рту. Напротив, выражение малоинтенсивной эмоции грусти приводит к возбуждению парасимпатической системы и проявляется в пониженном сердечном ритме и кровяном давлении. Но хотя интенсивность и помогает разграничивать эмоции грусти и радости, она не дает возможности разграничивать эмоции радости и гнева: для этой цели уже используется ось валентности.

Задача автоматической классификации эмоций в речи

Для системы, способной автоматически классифицировать эмоции в речи, необходимы две ее составные части: обучающие данные и алгоритм, который на основе этих обучающих данных будет способен классифицировать эмоции.

При сборе обучающих данных необходимо учесть множество параметров. Так, мужчины и женщины, а также пожилые и подростки по-разному выражают в своей речи эмоции, поэтому важно сбалансировать говорящих по полу и возрастным группам, чтобы классифицирующий алгоритм мог распознавать их с одинаковой уверенностью. Что касается самих данных, то, во-первых, классы эмоций должны быть сбалансированы, то есть иметь сходное количество примеров, а во-вторых, необходима выверенная длительность записанных высказываний, обычно в пределах 3–11 секунд [Lotfian, Busso 2019], чтобы в них было достаточно признаков для возможности классификации и не было наложения других эмоций (но при увеличении общего объема данных требования к длительности часто смягчаются). Наконец, важно учесть лексическое разнообразие высказываний, то есть количество в них уникальных слов: значение коэффициента TTR (Type/Token Ratio), отражающего лексическое разнообразие примеров, должно находиться в диапазоне 0,3–0,6, тем самым указывая на достаточную разнообразность лексики и в то же время допуская повторы, чтобы обучаемая модель имела способность к обобщению данных.

Обучение классификатора происходит на основе признаков, извлекаемых из обучающих данных. Так как мы рассматриваем эмоции в речи, то для начала нужно понять, какие именно акустические и лексические признаки коррелируют с указанными базовыми эмоциями и интенсивностью/валентностью. По общему представлению, речь, сопровождаемая высокоинтенсивными эмоциями, такими как

радость и гнев, обычно быстрая и энергичная, тогда как для малоинтенсивных эмоций, таких как грусть, напротив, характерна медленная и тихая, часто низко-тональная речь. Как основные признаки, которые часто используются в исследовании, здесь можно выделить высоту голоса, его громкость, длительность отдельных сегментов, а также значения формант, обычно первой и второй. Помимо этого, также часто используются признаки на основе амплитудного спектра — MFCC-признаки (англ. Mel-frequency cepstral coefficients) [Ali et al. 2021]. Для их расчета используется специальная *мел-шкала*, которая позволяет приблизить обработанный звук к тому, как его слышит человек, так как человеческое ухо более чувствительно к небольшим изменениям звука на низких частотах по сравнению с высокими. Все признаки обычно рассчитываются для каждого из *фреймов* — отрезков длиной 25 мс, на которые делят исходную аудиозапись.

Что же касается валентности, то здесь у исследователей нет четкого понимания того, с какими из акустических характеристик она коррелирует. Последние исследования показывают, что улучшить качество предсказания по шкале “валентность” помогает добавление к признакам имплицитной лингвистической информации [Wagner et al. 2023].

Данные для исследования

В качестве языков для исследования были взяты польский язык, родственный по отношению к русскому, китайский, имеющий сложную систему тонов, и японский. Для этих языков были соответственно взяты датасеты эмоциональной речи nEMO [Christop 2024], ESD [Zhou et al. 2022] и JNVN [Xin et al. 2024]. На их основе было проведено обучение моделей для задачи автоматического распознавания эмоций с целью их дальнейшего переноса и тестирования на данных русского датасета эмоциональной речи Dusha. Из-за разницы в количестве примеров и эмоций между разными датасетами для каждого иностранного языка был создана нормализованная укороченная версия соответствующего датасета размером в 600 высказываний, по 200 высказываний для каждой из трех эмоций: гнев, радость и грусть. Эти эмоции были выбраны на основе того, что они представлены во всех указанных выше датасетах.

Рассмотрим характерные признаки каждого из языков. Все эти языки так или иначе могут выражать эмоции через лексику (польск. «To cudownie!» — «Это чудесно!»), устойчивые словосочетания, интонацию и различные невербальные элементы, например смех. В японском часто используются частицы «не» и «уо», присоединяемые в конце предложения и выражающие восхищение, а также имеется развитая грамматическая система для выражения вежли-

ности, что потенциально может влиять на выражение эмоций. Помимо этого, в японском языке также присутствует музыкальное ударение, при котором высота звучания слога может быть смысло-различительной (в отличие от силового ударения в русском языке, которое только делает слог громче и/или удлиняет его). Китайский же язык имеет еще более сложную систему смысло-различительных тонов, которые могут накладываться на интонацию высказывания. В русском и польском языках более свободный порядок слов из-за наличия богатых падежных систем, хотя базовым по-прежнему остается порядок SVO, как и у китайского. У японского же, в отличие от других языков, базовым является порядок слов SOV.

В Таблице 1 приведены различия между русским и другими рассматриваемыми языками по основным лингвистическим признакам, рассчитанные на основе векторов lang2vec [Littell et al. 2017], применяемых для оценки косинусной близости между языками.

Таблица 1

Различия между русским и другими рассматриваемыми языками по основным лингвистическим признакам; чем значение ближе к нулю, тем выше сходство языков.

	Польский	Китайский	Японский
Синтаксическое различие	0.47	0.57	0.66
Фонологическое различие	0	0.62	0.43
Генетическое различие	0.4	1	1
Различие фонемного инвентаря	0.49	0.69	0.54

Данные о корпусах представлены в Таблице 2:

Таблица 2

Данные о датасетах, использованных в данной работе

Название датасета	Объем	Количество говорящих	Язык	Эмоции
Dusha	300000 высказываний	Более 2000	Русский	Гнев, Позитив, Грусть, Другое, Нейтральное состояние
nEMO	4481 высказываний	9	Польский	Гнев, Радость, Грусть, Страх, Удивление, Нейтральное состояние
ESD	17500 высказываний	10	Китайский	Гнев, Радость, Грусть, Удивление, Нейтральное состояние
JVNV	1615 высказываний	4	Японский	Гнев, Радость, Грусть, Отвращение, Страх, Удивление

Тестирование

В качестве основной речевой модели, использованной для исследования межъязыкового переноса моделей, была взята архитектура HuBERT-base [Hsu et al. 2021], содержащая около 95 млн параметров. При обучении она выполняет задачу классификации: маскирует часть поступающих на вход фреймов, которые затем учится классифицировать на основе конечного набора классов. Этот набор классов, в свою очередь, формируется с помощью применения к незамаскированным фреймам метода k-средних, который разбивает их множество на k кластеров. Для задачи классификации в модель был также добавлен конечный классифицирующий слой.

Для каждого иностранного датасета модель HuBERT была обучена на соответствующей выборке из 600 высказываний, из которых 20 % были использованы для валидации.

Общее качество классификаций каждой модели было подсчитано на основе метрики Unweighted Accuracy (UA), отражающей отношение количества правильно классифицированных примеров к неправильным. Для каждой эмоции было отдельно подсчитано качество ее классификации на основе F-меры. Полученные результаты представлены в Таблице 3.

Таблица 3

Результаты переноса на русский язык речевой модели HuBERT, обученной на соответствующем иностранном языке

Исходный язык	Качество классификации			
	anger-f1	happy-f1	sad-f1	UA
Польский	0.43	0.26	0.45	0.4
Китайский	0.1	0.17	0.52	0.37
Японский	0.4	0.38	0.45	0.41

Анализ результатов

Общее качество переноса моделей сходно для польского и японского языков, при этом китайская модель показывает более низкий результат. Среди эмоций при этом лучше всего классифицируется грусть, в то время как радость классифицируется хуже остальных. Это можно объяснить через пространство возбуждения/валентности. Выражение грусти обычно характеризуется низкой интенсивностью, в противовес гневу и радости, что позволяет модели более однозначно противопоставить грусть остальным эмоциям, так как для разной интенсивности свойственны характерные акустические признаки. Гнев и радость, однако, противопоставлены

между собой только по шкале валентности, эксплицитно не соотносящейся с акустическими признаками высказывания, которые могла бы использовать модель, из-за чего она большинство таких высказываний относит к эмоции гнева.

Японская модель не показывает сильного отличия от польской модели, за исключением немного более высокого качества классификации эмоции радости. Что касается распределения классифицированных моделями высказываний по классам, то у японской модели оно получилось самым равномерным (около 33 % тестовых высказываний в каждом классе) из-за чего f -мера более высокая, тогда как польская модель относит около 50 % всех высказываний к эмоции гнева. Каких-либо значимых различий из-за порядка слов SOV в японском не наблюдается: например, высказывание «а мне нравится ход ваших мыслей», произнесенное с грустной интонацией, и польской, и японской моделью неверно классифицируется с эмоцией гнева.

На общем фоне выделяется китайская модель, имеющая наилучшее качество классификации эмоции грусти, но наихудшее у других эмоций. Распределение классифицированных высказываний у китайской модели еще более неравномерное, и она относит около 90 % всех тестовых высказываний к эмоции грусти, что, вероятнее всего, указывает на влияние китайской тональной системы на классификацию.

Таким образом, полученные результаты указывают на большее влияние интонации и акустических характеристик при классификации эмоций в речи, если сравнивать их с синтаксисом и лексикой. В качестве потенциального дальнейшего исследования может выступить изучение мультимодальной модели, имеющей лексическую информацию непосредственно из расшифровки аудиозаписей.

Заключение

В данной статье было рассмотрено влияние лингвистических характеристик, свойственных конкретным языкам, на поведение речевых моделей, обученных на этих языках с целью решения задачи автоматического распознавания эмоций в речи. Был проведен эксперимент с обучением модели HuBERT на польском, китайском и японском языках, с дальнейшим переносом обученной модели на русский язык. Результаты показали, что на множестве рассмотренных языков наибольшее влияние на качество классификации при переносе модели на другой язык оказывают просодические характеристики языка-источника, в то время как различия в синтаксисе практически не оказывают влияния на способность модели распознавать эмоции. Полученные данные могут быть полезны при

решении задачи распознавания эмоций в речи говорящих на польском, китайском и японском как родном, а также изучающих русский как иностранный.

СПИСОК ЛИТЕРАТУРЫ

1. *Верхоляк О.В.* Автоматическое распознавание эмоциональных состояний дикторов по голосовым характеристикам и тональности текста высказывания. Диссертация на соискание ученой степени кандидата технических наук. Санкт-Петербург. 2001. URL: http://fppo.ifmo.ru/?page1=16&page2=86&number_file=D FC2E78CE850F9B8BAA24619873810B3.
2. *Ali S., Tanweer S., Khalid S., Rao N.* Mel Frequency Cepstral Coefficient: A Review // ICIDSSD. 2021.
3. *Christop I.* nEMO: Dataset of Emotional Speech in Polish // Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024). Torino, Italia. 2024. P. 12111–12116.
4. *Cowie R., Douglas-Cowie E., Tsapatsoulis N., Votsis G., Kollias S. et. al.* Emotion recognition in human-computer interaction // IEEE Signal Processing Magazine. 2001. Vol. 18. № 1. P. 32–80.
5. *Hansen L., Zhang Y.P., Wolf D., Sechidis K., Ladegaard N. et. al.* A generalizable speech emotion recognition model reveals depression and remission // Acta Psychiatrica Scandinavica. 2022. Vol. 145. № 2. P. 186–199.
6. *Hsu W.N., Bolte B., Tsai Y.H., Lakhota K., Salakhutdinov R. et. al.* HuBERT: Self-Supervised Speech Representation Learning by Masked Prediction of Hidden Units // IEEE/ACM Transactions on Audio, Speech, and Language Processing. 2021. Vol. 29. P. 3451–3460.
7. *Jackson J.K., Watts J., Henry T.R., List J.-M., Forkel R. et. al.* Emotion Semantics show both cultural variation and universal structure // Science. 2019. Vol. 366. № 6472. P. 1517–1522.
8. *Kondratenko V., Sokolov A., Karpov N., Kutuzov O., Savushkin N. et. al.* Large Raw Emotional Dataset with Aggregation Mechanism // arXiv.org. 2022. URL: <https://arxiv.org/abs/2212.12266> (дата обращения: 20.02.2025).
9. *Lefter I., Jonker C.M.* Aggression recognition using overlapping speech // 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII). San Antonio, TX, USA. 2017. P. 299–304.
10. *Lindquist K.A.* Language and Emotion: Introduction to the Special Issue // Affective Science. 2021. Vol. 2. P. 91–98.
11. *Littell P., Mortensen D.R., Lin K., Kairis K., Turner C. et. al.* URIEL and lang2vec: Representing languages as typological, geographical, and phylogenetic vectors // Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics. Valencia, Spain. 2017. Vol. 2. P. 8–14.
12. *Lotfian R., Busso C.* Building Naturalistic Emotionally Balanced Speech Corpus by Retrieving Emotional Speech from Existing Podcast Recordings // IEEE Transactions on Affective Computing. 2019. Vol. 10. № 4. P. 471–483.
13. *Sboev A., Naumov A., Rybka R.* Data-Driven Model for Emotion Detection in Russian Texts // Procedia Computer Science. 2021. Vol. 190. P. 637–642.
14. *Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L. et. al.* Attention Is All You Need // arXiv.org. 2017. URL: <https://arxiv.org/abs/1706.03762> (дата обращения: 24.02.2025).
15. *Wagner J., Triantafyllopoulos A., Wierstorf H., Schmitt M., Burkhardt F. et. al.* Dawn of the Transformer Era in Speech Emotion Recognition: Closing the Valence Gap //

IEEE Transactions on Pattern Analysis and Machine Intelligence. 2023. Vol. 45. № 9. P. 10745–10759.

16. Xin D., Jiang J., Takamichi S., Saito Y., Aizawa A. et. al. JVN-V: A Corpus of Japanese Emotional Speech with Verbal Content and Nonverbal Expressions // IEEE Access. 2024. Vol. 12. P. 19752–19764.
17. Zhou K., Sisman B., Liu R., Li H. Emotional voice conversion: Theory, databases and ESD // Speech Communication. 2022. Vol. 137. P. 1–18.

REFERENCES

1. Verkholyak O. Automatic Recognition of Speaker's Emotional States Based on Audio and Text. Academic dissertation candidate of engineering. Saint Petersburg, 2021. URL: http://fppo.ifmo.ru/?page1=16&page2=86&number_file=DFC2E78CE850F9B8BAA24619873810B3.
2. Ali S., Tanweer S., Khalid S., & Rao N. Mel Frequency Cepstral Coefficient: A Review. *ICIDSSD*. 2021. doi: 10.4108/eai.27-2-2020.2303173.
3. Christop I. nEMO: Dataset of Emotional Speech in Polish. *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*. Torino, Italia, ELRA and ECCL, 2024, pp. 12111–12116.
4. Cowie R., Douglas-Cowie E., Tsapatsoulis N., Votsis G., Kollias S. et. al. G. Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine*. 2001, vol. 18, № 1, pp. 32–80. doi: 10.1109/79.911197.
5. Hansen L., Zhang Y.P., Wolf D., Sechidis K., Ladegaard N. et. al. A generalizable speech emotion recognition model reveals depression and remission. *Acta Psychiatrica Scandinavica*. 2022, vol. 145, № 2, pp. 186–199. doi: 10.1111/acps.13388.
6. Hsu W.N., Bolte B., Tsai Y.H., Lakhotia K., Salakhutdinov R. et. al. HuBERT: Self-Supervised Speech Representation Learning by Masked Prediction of Hidden Units. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. IEEE Press, 2021, vol. 29, pp. 3451–3460. doi: 10.1109/TASLP.2021.3122291.
7. Jackson J.K., Watts J., Henry T.R., List J.-M., Forkel R. et. al. Emotion Semantics show both cultural variation and universal structure. *Science*. 2019, Vol. 366, № 6472, pp. 1517–1522. doi: 10.1126/science.aaw8160.
8. Kondratenko V., Sokolov A., Karpov N., Kutuzov O., Savushkin N. et. al. Large Raw Emotional Dataset with Aggregation Mechanism. *arXiv.org*. 2022. URL: <https://arxiv.org/abs/2212.12266> (accessed: 20.02.2025).
9. Lefter I., Jonker C.M. Aggression recognition using overlapping speech. *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. San Antonio, TX, USA, IEEE, 2017, pp. 299–304. doi: 10.1109/ACII.2017.8273616.
10. Lindquist K.A. Language and Emotion: Introduction to the Special Issue. *Affective Science*. 2021, vol. 2, pp. 91–98. doi: 10.1007/s42761-021-00049-7.
11. Littell P., Mortensen D.R., Lin K., Kairis K., Turner C. et. al. URIEL and lang2vec: Representing languages as typological, geographical, and phylogenetic vectors. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*. Valencia, Spain, Association for Computational Linguistics, 2017, vol. 2, pp. 8–14. doi: 10.18653/v1/E17-2002.
12. Lotfian R., Busso C. Building Naturalistic Emotionally Balanced Speech Corpus by Retrieving Emotional Speech from Existing Podcast Recordings. *IEEE Transactions on Affective Computing*. 2019, vol. 10, № 4, pp. 471–483. doi: 10.1109/TAFCC.2017.2736999.

13. Sboev A., Naumov A., Rybka R. Data-Driven Model for Emotion Detection in Russian Texts. *Procedia Computer Science*. 2021, vol. 190, pp. 637–642. doi: 10.1016/j.procs.2021.06.075.
14. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L. et. al. Attention Is All You Need. *arXiv.org*. 2017. URL: <https://arxiv.org/abs/1706.03762> (accessed: 24.02.2025).
15. Wagner J., Triantafyllopoulos A., Wierstorf H., Schmitt M., Burkhardt F. et. al. Dawn of the Transformer Era in Speech Emotion Recognition: Closing the Valence Gap. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2023, vol. 45, № 9, pp. 10745–10759. doi: 10.1109/TPAMI.2023.3263585.
16. Xin D., Jiang J., Takamichi S., Saito Y., Aizawa A. et. al. JNVN: A Corpus of Japanese Emotional Speech with Verbal Content and Nonverbal Expressions. *IEEE Access*. IEEE, 2024, vol. 12, pp. 19752–19764. doi: 10.1109/ACCESS.2024.3360885.
17. Zhou K., Sisman B., Liu R., Li H. Emotional voice conversion: Theory, databases and ESD. *Speech Communication*. 2022, vol. 137, pp. 1–18. doi: 10.1016/j.specom.2021.11.006.

Поступила в редакцию 23.03.2025

Принята к публикации 17.06.2025

Отредактирована 20.06.2025

Received 23.03.2025

Accepted 17.06.2025

Revised 20.06.2025

ОБ АВТОРАХ

Лемаев Владислав Игоревич — аспирант кафедры теоретической и прикладной лингвистики филологического факультета МГУ имени М.В. Ломоносова; vladzhkv98@mail.ru

Лукашевич Наталья Валентиновна — доктор технических наук, кандидат физико-математических наук, профессор кафедры теоретической и прикладной лингвистики филологического факультета МГУ имени М.В. Ломоносова; ведущий научный сотрудник НИВЦ МГУ; louk_nat@mail.ru

ABOUT THE AUTHORS

Vladislav Lemaev — PhD student, Department of Theoretical and Applied Linguistics, Faculty of Philology, Lomonosov Moscow State University; vladzhkv98@mail.ru

Natalia Loukachevitch — Prof. Dr., Department of Theoretical and Applied Linguistics, Faculty of Philology, Lomonosov Moscow State University; louk_nat@mail.ru