

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ

САНКТ-ПЕТЕРБУРГСКИЙ ФЕДЕРАЛЬНЫЙ
ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР
РОССИЙСКОЙ АКАДЕМИИ НАУК

**Десятый
междисциплинарный семинар
«Анализ разговорной русской речи»**

АР³-2023

29–30 июня 2023 года,
Санкт-Петербург, СПбГУ

Санкт-Петербург

2023

ББК 32.965+81.1

А64

Научные редакторы:
У. Е. Кочеткова, П. А. Скрелин

А64 Анализ разговорной русской речи (АР³-2023): Труды десятого междисциплинарного семинара. — СПб.: Скифия-принт, 2023. — 90 с.

ISBN 978-5-00197-081-1

Издание представляет собой сборник докладов, сделанных на заседаниях десятого междисциплинарного семинара «Анализ разговорной русской речи» (АР³-2023), проходившего 29–30 июня 2023 года в Санкт-Петербургском государственном университете. Тематика докладов определялась (но не ограничивалась) анализом и моделированием различных видов речи на материале русского и других языков. Статьи сопровождаются аннотациями на английском языке.

ББК 32.965+81.1

*Издание осуществлено при поддержке
Special Interest Group on Russian Speech Analysis of
International Speech Communication Association*

ISBN 978-5-00197-081-1

© Коллектив авторов, 2023

ОГЛАВЛЕНИЕ

<i>Н. В. Богданова-Бегларян</i> Финальная вокализация в устном дискурсе как прием речевого манипулирования (эффект обманутого ожидания).....	5
<i>Д. В. Гребенкин</i> Использование современных сетей Хопфилда для обучения нейросетевых алгоритмов распознавания речи на неразмеченных данных.....	10
<i>В. В. Евдокимова, М. Р. Максимова</i> Сравнение формантных траекторий гласных в различных языках.....	16
<i>О. В. Камрыш</i> Универсальные и уникальные черты в интонации французских вопросительных высказываний.....	22
<i>Г. Е. Кедрова, М. В. Волкова, Н. В. Анисимов</i> Сопоставительный анализ дыхательной активности диктора в состоянии покоя и при чтении прозаического и поэтического текстов	28
<i>У. Е. Кочеткова, Т. В. Качковская, А. А. Соболева</i> Интонация завершенности и незавершенности в русской речи бразильцев	34
<i>Е. Е. Ляксо, О. В. Фролова, А. С. Николаев</i> Анализ эмоциональной речи детей	40
<i>А. А. Поволоцкая, А. А. Карпов</i> Современные проблемы автоматического распознавания речи: выявление и анализ экстралингвистических вокализаций в спонтанной разговорной речи	44
<i>Д. Н. Подвигина</i> О нейронной основе шкал оценки коротких интервалов времени.....	51
<i>П. А. Скредин, У. Е. Кочеткова, В. В. Евдокимова, Д. Д. Новосёлова</i> Анализ через синтез: установление акустических ключей иронии с помощью модификаций звукового сигнала	55
<i>Я. Сян, Н. В. Богданова-Бегларян</i> Аппроксимация и хеджирование в языке и речи	60

<i>Е. В. Фаненштль</i>	
Автоматическая кластеризация тональных акцентов в русском языке.....	66
<i>П. А. Холявин</i>	
Сравнение мелодических контуров односложных синтагм в русском и английском языках	72
<i>Г. Чжоу, С. О. Тананайко</i>	
The Second Formant Dynamics in Russian and Chinese Syllables	78
Аннотации докладов, сделанных на семинаре «Анализ разговорной русской речи» AP³-2023 (на английском языке)	
Abstracts of papers presented at workshop “Analysis of the Russian colloquial speech 2023” (in English).....	82

ФИНАЛЬНАЯ ВОКАЛИЗАЦИЯ В УСТНОМ ДИСКУРСЕ КАК ПРИЕМ РЕЧЕВОГО МАНИПУЛИРОВАНИЯ (ЭФФЕКТ ОБМАНУТОГО ОЖИДАНИЯ)¹

Наталья Викторовна Богданова-Бегларян
(Санкт-Петербургский государственный университет)

Устная спонтанная речь (СР) – это результат линейного процесса обдумывания и порождения текста, исследователи неизменно подчеркивают свободу СР, которой свойственны ошибки, неполнота, самоперебивы, самокоррекция, хезитации и пр. [9: 4–5]. Все эти особенности можно объединить понятием *хезитационные явления* (ХЯ), они не раз становились объектом специальных исследований, в том числе на неродном для говорящего языке (см., например: [8; 10]).

В настоящем исследовании в центре внимания находятся *вокализации* как разновидность ХЯ (*м-м, э-э*). Это один из способов невербального заполнения паузы хезитации (ПХ), под ними понимают «неречевые», или «речеподобные», звуки, или звуки «нефонемного характера». Такие элементы в целом рассматриваются как форма *речевого сбоя*, при котором нарушается плавное развертывание речевого потока, это разновидность «перерыва, который говорящий использует для подготовки следующей порции и/или (при сочетании с коррекцией) — для обдумывания возможного способа исправления предшествующей порции» [5].

Вокализации как ХЯ широко распространены в спонтанной речи на любом языке и способны вызвать затруднения при общении людей. Однако возможен и иной взгляд на это явление. Так, исследования показали, что из трех способов реализации в устной речи пауз хезитации наиболее неудачным, способным привести к коммуникативному сбою, является физическая пауза (молчание, отсутствие фонации). Вербальное заполнение ПХ так называемыми «словами-паразитами» (*это, ну, это самое, как его (её, их)* и под.) свидетельствует о низком уровне речевой культуры (компетенции) говорящего [4: 91]. Высокому же уровню свойственно использование в этой роли именно разного рода вокализаций [3]. С этой точки зрения вокализации могут расцениваться исключительно положительно и даже рассматриваться как своеобразный прием *речевого*

¹ Исследование проведено при финансовой поддержке гранта Санкт-Петербургского государственного университета (проект № 94033528 «Моделирование коммуникативного поведения жителей российского мегаполиса в социально-речевом и прагматическом аспектах с привлечением методов искусственного интеллекта»).

манипулирования: с их помощью говорящий не только заполняет паузу колебания и дает себе время на поиск нужного слова, но и может удерживать внимание собеседника и не допускать его вступления в диалог до окончания своей реплики.

Речевое манипулирование лишь в конце XX века стало объектом активного внимания *когнитивной лингвистики*, прежде всего при описании политического, рекламного, медиа-дискурса и под., и в таких прикладных областях, как предвыборные кампании, идеологическая пропаганда, печатные и телевизионные СМИ, обучение, психотерапия. На материале повседневной речи оно рассматривалось в работе [1], посвященной таким единицам, как контактные глаголы (*знаешь/те, прикинь, слушай/те* и под.); конструкции-коллокации, созданные на их основе (*знаешь как...; не видишь что ли?!; разговорные конструкции типа с ума сошёл что ли?! дурак что ли?!; а также маркеры да? и ну?* Все они объединяются в класс неясных *манипулятивных единиц* (МЕ) нашей бытовой коммуникации, прежде всего в силу неосознаваемости говорящим и речевого автоматизма при порождении. С помощью МЕ говорящий «словно бы держит собеседника “на поводке”, манипулирует им, как марионеткой, не дает ему отвлечься и все время проверяет наличие внимания с его стороны» [1: 291].

Можно было бы ожидать, что такую же роль играют в диалоге конечные (в рамках реплики) хезитации-вокализации, позволяющие говорящему удержать за собой очередь говорения и не допускающие захвата этой очереди (*turn-taking*) собеседником (о «коммуникативных девайсах» см.: [7]). По крайней мере, именно по такой способности вокализации гипотетически противостоят физическим ПХ, которые как раз маркируют отсутствие запрета говорить.

Между тем, анализ корпусного материала (УП НКРЯ и ОРД) этих гипотетических ожиданий не оправдал. Сплошь и рядом, особенно в общении близких людей, вокализации не справляются с ролью манипулятивных единиц, ср.:

- 1) **Ж1** ну мы и сухого / **знаешь** / больше ста пятидесяти не выпиваем ... # **И4** мы в общем тоже не то что бы (э-э) ... @ **Ж2** конечно // @ **Ж1** нет / ну сладкого мы вообще ... @ **И4** литрами (ОРД)1;
- 2) **Ж12** я сегодня себе ногу повредила // # **И5** да ты что ! ногу / повредила // а чем ? # **Ж12** да ... мне вот посюда // ***П** у меня (э-э) # И5 на физкультуре ? # ***П** ну-у / (...) не совсем конечно // ***П** меня что-то мячом попали (ОРД);

¹ Знак # в расшифровках ОРД означает мену говорящих, знак @ — наложение их речи (одновременное говорение). О других особенностях орфографического представления материала ОРД см. подробнее: [6: 242-243].

- 3) **Ж1** а они знаешь / они их там в холодильнике держат / они все (э-э) # **Ж2** а Галька% не приехала / никто не звонил им туда ? *П # **Ж1** я не звонила / думаю / что она не приехала ещё (ОРД).

В приведенных контекстах собеседники, не обращая внимания на вокализации, иногда хором, перебивают основного говорящего, а он, кажется, нисколько на это не обижается и продолжает начатую мысль или переключается на заданный вопрос (все его реплики в примерах подчеркнуты).

Так, в примере (1) информант И4 начинает фразу: мы в общем тоже не то что бы (э-э) ... — и после вмешательства сразу двух собеседников (**Ж2** и **Ж1**), говорящих одновременно и не воспринявших «предупреждающей» вокализации, спокойно заканчивает ее: литрами.

Похожая ситуация и в примере (2): коммуникант-женщина **Ж12**, разговаривающая с информантом И5, начинает рассказ о том, как она повредила ногу (*я сегодня себе ногу повредила*), коротко реагирует подтверждающим *да* на эмоциональную реплику собеседника (*да ты что ! ногу / повредила // а чем ?*) и продолжает говорить: *мне вот посюда // *П у меня (э-э)*. Надо отметить, что после второго перебивания ее информантом (*на физкультуре ?*) **Ж12** уже не реагирует на заданный вопрос, а уточняет свою изначальную информацию о том, что нога именно повреждена: *ну-у / (...) не совсем конечно // *П в меня что-то мячом попали*. Вокализация снова не смогла остановить собеседника от вмешательства в рассказ.

В контексте (3) говорящая **Ж1** сбилась с мысли, начала хезитировать, не закончив фразы, — и собеседница **Ж2** тут же вступила в разговор, задав вопрос, никак не связанный с тем, о чем только что шла речь: *а Галька% не приехала / никто не звонил им туда ?* **Ж1** покорно приняла такой поворот разговора, не стала возвращаться к своей теме и ответила на заданные вопросы: *я не звонила / думаю / что она не приехала ещё*. Финальная вокализация снова не стала средством манипуляции (даже если **Ж1** и вкладывала в нее подсознательно такой смысл, о чем можно, впрочем, только догадываться), а лишь маркировала речевой сбой и замешательство первой говорящей.

Встретились, впрочем, и контексты, когда функция вокализации была правильно воспринята собеседником, он выждал со своей репликой (о чем свидетельствует наличие паузы (*П) после вокализации), но продолжения не дождался — и перевел очередь говорения на себя (все реплики основного говорящего, а также предупреждающие реплики с вокализацией снова подчеркнуты):

- 4) **И102** а / разве не надо ничего резать // *П # **Ж1** нет // # **И102** а просто я думал / я же забыл про него / давай / конечно / чего там

// # Ж1 ой / *Н / ты *Н не забыл // *П И102 э-э *П Ж1 ну / для порядка (ОРД);

- 5) И106 (э-э) у меня сегодня только виртуальная шляпа / а концерт ещё и с бесплатным входом был / то есть (э-э) ... *П М1 Ася% это какая? @ И106 ну / а ей бы вообще ничего не перепало / я говорю ладно / держи сто рублей (ОРД).

Функция МЕ вокализацией и здесь оказалась не реализована. Так, в примере (5) говорящий И106, которого собеседник М1 перебил вопросом (*Ася% это какая?*), не стал ни отвечать на этот вопрос, ни продолжать начатую фразу, а просто бросил ее и произнес другую (*ну / а ей бы вообще ничего не перепало / я говорю ладно / держи сто рублей*).

Любопытным оказался следующий контекст (6):

- 6) И4 я думаю / что он на порядок дороже / *П и я думаю / что вот (э-э) @ Ж1*Н *П # И4 тот (...) тот в такой очень средней записи держался пять часов / а (э-э) @ Р1 *Н *П # Ж1 а *Н # И4 а на очень плохой семнадцать // но (э-э) это была настолько плохая / что просто не разобрать ничего / вот это *П / просто ужас какой-то # Ж1 а (э-э) что за испытания такие идут? (ОРД).

В разговоре о диктофоне, используемом для записи корпусного материала, И4 довольно жестко удерживает за собой очередь говорения и последовательно маркирует возникающие заминки вокализацией э-э. Однако его собеседники — женщина Ж1 и ребенок Р1 — после каждой такой вокализации пытаются вступить в разговор (знак *Н означает неразборчивый фрагмент), но И4 «подавляет» все такие попытки и упорно продолжает свое высказывание. Так или иначе, и в этом случае функция манипулятивного средства, отведенная говорящим вокализации, осталась нереализованной.

Ожидаемых контекстов, когда финальное э-э удержало бы собеседника от вступления в разговор, в рассмотренном корпусном материале попросту не встретилось.

Зато обнаружили примеры, когда вокализация воспринимается собеседником как призыв помочь говорящему в подборе слова — и тогда фраза выстраивается по принципу *backchannel*¹:

- 7) Лена Да/ но мы с ней немного/ э-э... Леонид Поссорились? Лена Угу (УП).

¹ Подробнее о таком — более широком по сравнению с традиционной «обратной связью» [13] — понимании явления *backchannel*, когда высказывание строится не одним, а сразу несколькими говорящими, см.: [11; 2; 12].

По-видимому, детальный анализ корпусного материала может выявить еще какие-то функции вокализации, кроме описанных здесь и в научной литературе, но относить вокализации к средствам речевого манипулирования, похоже, нет серьезных оснований.

Список использованной литературы

1. *Богданова-Бегларян Н. В.* О неявных средствах манипуляции в повседневном дискурсе // Речевое воздействие в разных дискурсах. Т. 5 / под ред. Ж. Сладкевич, А. Климкевич, К. Велондек. — Gdańsk: Wydawnictwo Uniwersytetu Gdańskiego, 2021. — С. 289–298.
2. *Гренобль Л. А.* Синтаксис и совместное построение в устном русском диалоге // Вопросы языкознания. — № 1, 2008. — С. 25–36.
3. Звуковой корпус как материал для анализа русской речи. Коллективная монография. Часть 1. Чтение. Пересказ. Описание / отв. ред. Н. В. Богданова-Бегларян. — СПб.: Филологический ф-т СПбГУ, 2013. — 532 с.
4. *Земская Е. А.* Русская разговорная речь: лингвистический анализ и проблемы обучения. 2-е изд. — М.: Русский язык, 1987. — 237 с.
5. *Подлеская В. И., Кибрик А. А.* Коррекция сбоев в устной спонтанной речи: опыт корпусного исследования // Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции «Диалог’05» (Звенигород, 17 июля 2005 г.) / под ред. И. М. Кобозевой, А. С. Нариньяни, В. П. Селегея. — М.: РГГУ, 2005 [Электронный ресурс]. <http://www.dialog-21.ru/media/2416/podlesskaya-kibrik.pdf> (2005).
6. Русский язык повседневного общения: особенности функционирования в разных социальных группах. Коллективная монография / под ред. Н. В. Богдановой-Бегларян. — СПб.: ЛАЙКА, 2016. — 244 с.
7. *Сакс Х., Джефферсон Г., Щеглофф Э. А.* Простейшая систематика организации очередности в разговоре // Социологический журнал. — Т. 14, № 1, 2015. — С. 142–202.
8. *Тэн Хай.* Универсальные и типологические черты паузации в спонтанной речи носителей разных языков // Теоретическая и прикладная лингвистика. — 2015. Том 1, № 2. — С. 105–113.
9. Фонетика спонтанной речи / под ред. Н. Д. Светозаровой. — Л.: Изд-во ЛГУ, 1988. — 243 с.
10. *Чэн Чэнь.* Хезитации в русской устной речи носителей китайского языка. — СПб.: Нестор-История, 2021. — 232 с.
11. *Cathcart N., Carletta J., Klein E.* A Shallow Model of Backchannel Continuers in Spoken Dialogue // 10th Conference of the European Chapter of the ACL. — Vol. 1, 2003. — Pp. 51–58.
12. *Grenoble L.A.* Talking out of Turn: (Co)-Construction Russian Conversation // Approaches to Slavic Interaction / eds N. Thielemann, P. Costa. — Amsterdam: John Benjamins, 2013. — Pp. 17–33.
13. *Yngve V. H.* On Getting a Word in Edgewise. Papers from the Sixth Regional Meeting of the Chicago Linguistic Society. — Chicago, Illinois: University of Chicago, Department of Linguistics, 1970. — Pp. 567–578.

ИСПОЛЬЗОВАНИЕ СОВРЕМЕННЫХ СЕТЕЙ ХОПФИЛДА ДЛЯ ОБУЧЕНИЯ НЕЙРОСЕТЕВЫХ АЛГОРИТМОВ РАСПОЗНАВАНИЯ РЕЧИ НА НЕРАЗМЕЧЕННЫХ ДАННЫХ

*Даниил Витальевич Гребенкин
(Новосибирский государственный университет)*

Сегодня алгоритмы распознавания речи применяются во многих электронных устройствах, которыми человек пользуется каждый день — голосовые помощники, аудиопереводчики, «умные» телевизоры и др. Для создания современных систем распознавания речи широко применяется интегральный (end-to-end) подход [1] из-за ряда преимуществ: для интегральных моделей нет необходимости в отдельном формировании словаря и обучения языковой модели; методы квантизации и прунинга позволяют уменьшить размер нейросетей, увеличить скорость их работы; с помощью переноса обучения [2] и небольшого количества размеченных аудиозаписей возможна адаптация больших заранее обученных моделей к новой предметной области. Наибольшую популярность интегральный подход получил после появления архитектуры «трансформер» в 2017 году, описанной в работе [3]. Трансформер состоит из стеков декодировщиков и кодировщиков, которые связаны между собой:

- 1) каждый из кодировщиков в стеке состоит из слоя внутреннего внимания и нейронной сети прямого распространения;
- 2) каждый из декодировщиков в стеке включает в себя слой внутреннего внимания, нейронную сеть прямого распространения и слой внимания (encoder-decoder attention) между ними.

В основе архитектуры трансформеров лежит так называемый «механизм внимания». Обобщенный механизм внимания был предназначен для использования в рекуррентных и сверточных нейронных сетях для нахождения взаимосвязей во входной и выходной последовательности символов в задачах машинного перевода [4]. Дальнейшим развитием стал «внутренний механизм внимания» («self-attention»), который позволял найти различия между словами входной последовательности в течение ее обработки кодировщиком. Он использовался и во всех последующих работах, развивавших оригинальную концепцию трансформера для применения в задачах распознавания речи: внедрение сверточных нейронных сетей для уменьшения размерности между входами и выходами модели [5], использование кодировщиков и новой функции потерь для учета уже обработанного контекста [6], сочетание преимуществ

трансформера и сверточных нейронных сетей [7] для учета локального и глобального контекста [8], самообучение («self-supervised learning») на неразмеченных речевых данных для дальнейшего превращения обученных речевых представлений в слова естественного языка [9].

Однако, в ходе использования трансформеров исследователями были обнаружены некоторые недостатки механизма внимания. В работе [10] авторы показали, что модели на базе трансформеров склонны к переобучению на обучающей выборке, состоящей из последовательностей определенной длины. Было доказано, что такие модели работают хуже на более коротких или на очень длинных последовательностях на этапе тестирования. Авторы работы [11] исследовали способность трансформеров распознавать различные классы формальных языков. Результаты этих экспериментов говорят о том, что все языки «со звездочками» («non-star-free») и некоторые языки «без звездочек» («star-free») не были распознаны моделями с механизмом внимания.

В качестве одного из аналогов внутреннего механизма внимания в исследовании [12] была предложена ассоциативная память. Само понятие «ассоциативной памяти» основано на представлениях о человеческой памяти: некоторые объекты могут вызывать фрагменты из прошлого, которые ассоциированы с этим объектом. Выделяют два типа моделей ассоциативной памяти на нейронных сетях [13]:

1. Автоассоциативная память может завершить или исправить входной образ, но не может ассоциировать полученный образ с другим образом;
2. Гетероассоциативная память входному набору образов ставит в соответствие выходной (сети Хэмминга).

Архитектура нейросети, моделирующая автоассоциативную память, была впервые предложена в 1982 году в работе [14] Джоном Хопфилдом, нейросеть была способна исправить или дополнить полученный на вход «образ». Главной проблемой такой модели являлось то, что количество «образов», на которые могла быть обучена такая модель, было ограничено. В 2016 году в исследовании [15] было показано, что это количество может быть существенно увеличено с изменением функции энергии, описывающей работу нейросети. Такая модификация получила названия «плотной ассоциативной памяти» (dense associative memory (DAM)) и «современных сетей Хопфилда». В Python-библиотеке *hflayers* авторами [16] были реализованы слои DAM для их применения в архитектурах на основе трансформера.

В данном исследовании выдвигается гипотеза о том, что использование «плотной» ассоциативной памяти вместо механизма внимания

позволит сделать модель распознавания речи устойчивой к шумам, реверберационным эффектам и особенностям речи (диалектам, нарушениям произношения), поскольку такая модифицированная модель будет содержать обученные речевые представления (образы), которые могут быть восстановлены из любой последовательности входных данных с помощью ассоциативной памяти. Для проверки этой гипотезы было создано две модели на основе нейросетевого алгоритма Wav2Vec2: с механизмом внимания (*w2v2-classic-tiny*) и с имплементациями современных сетей Хопфилда из работы [16] (*w2v2-hopfield-tiny*). Обучение обеих моделей проводилось на корпусах SOVA RuYoutube (более 2000 часов речи на русском языке) и SOVA RuDevices (100 часов) [17], тестирование — на *voxforge-ru* [18] и Russian Golos [19], использовалась метрика качества, определяющая количество неверно распознанных символов (Character Error Rate; CER) (табл. 1). Для тестирования устойчивости моделей к звуковым артефактам и эффектам аудиозаписи корпуса *voxforge-ru* был замиксованы с несколькими типами шумов и различными значениями отношения сигнал/шум (ОСШ) с помощью разработанной библиотеки [20]. Результаты распознавания моделей модифицированных данных представлены в таблице 2.

Таблица 1

Значения CER (%) на тестовых данных

Тестовый корпус	<i>w2v2-classic-tiny</i>	<i>w2v2-hopfield-tiny</i>
<i>voxforge-ru</i>	47.6	<u>35.9</u>
<i>sberdevices-golos-crowd (test)</i>	48.9	<u>35.5</u>
<i>sberdevices-golos-farfield (test)</i>	57.4	<u>45.6</u>

Таблица 2

Значения CER (%) на модифицированном *voxforge-ru*

Тип шума	ОСШ (в дБ)	<i>w2v2-classic-tiny</i>	<i>w2v2-hopfield-tiny</i>
Звуки домашних животных	5.0	<u>54.9</u>	55.0
Звуки домашних животных	10.0	48.4	48.4
Звуки домашних животных	15.0	43.8	43.8
Речевой фоновый шум	5.0	63.9	<u>63.8</u>
Речевой фоновый шум	10.0	<u>54.8</u>	54.9
Речевой фоновый шум	15.0	47.8	<u>47.6</u>

Для того, чтобы объяснить полученные результаты, был выполнен анализ различных классов звуков, предсказанных моделью, с использованием принудительного выравнивания («forced alignment») [21] по следующему алгоритму:

- 1) Обработка аудио новой моделью (*w2v2-hopfield-tiny* или *w2v2-classic-tiny*) и извлечение выходов последнего слоя новой модели;
- 2) Обработка аудио моделью распознавания речи [22] и нахождение временных границ каждого звука, который в транскрипции представлен буквой;
- 3) Усреднение найденных по границам массивов векторных представлений.

Оценка эффективности (компактности) получившихся речевых векторных представлений проводилась с помощью коэффициента Силуэта, который показывает, насколько каждый объект (вектор) «похож» на другие объекты в том классе речевых представлений, в который он был распределен в процессе кластеризации, и «не похож» на объекты из других классов (табл. 3).

Таблица 3

Значения коэффициента Силуэта для объектов классов векторных представлений. Объекты, имеющие значения близкие к 1, наиболее похожи на другие объекты в том же классе; объекты, имеющие значения близкие к 0, расположены на границе классов

Буквенные представления фонем	<i>w2v2-classic-tiny</i>	<i>w2v2-hopfield-tiny</i>
ударные гласные	0.0613	<u>0.0615</u>
безударные гласные	0.0048	<u>0.0050</u>
сонорные согласные ('р', 'л', 'м', 'н')	<u>0.1469</u>	0.1365
смычные согласные ('б', 'п', 'д', 'т', 'г', 'к')	0.1059	<u>0.1073</u>
аффрикаты ('ч', 'ц')	0.2458	<u>0.2763</u>
щелевые согласные ('ф', 'в', 'с', 'з', 'ш', 'ж', 'х')	<u>0.0842</u>	0.0788

В результате исследования удалось создать модифицированную архитектуру модели Wav2Vec2 со слоями «плотной» ассоциативной памяти и обучить новую модель на неразмеченных речевых данных. Результаты тестирования на речевых корпусах показали, что применение слоев «плотной» ассоциативной памяти позволило снизить CER для

модели, обученной с использованием небольшого количества тренировочных аудио. Лучшая компактность речевых векторных представлений модели с новой архитектурой была продемонстрирована только на фонетических классах с относительно большой длительностью (гласные). Дальнейшая работа будет сосредоточена на двух направлениях: обучение моделей с разными архитектурами на большем количестве неразмеченных речевых данных для подтверждения/опровержения гипотезы эффективности работы модели на фонетических классах с большей длительностью; обучение мультязычной модели с слоями Хопфилда для изучения моделирования ассоциативных зависимостей между фонемами разных языков.

Список использованной литературы

1. Марковников Н. М., Куняткова И. С. Аналитический обзор интегральных систем распознавания речи // Тр. СПИИРАН. — 2018. — Т. 3, № 58. — С. 77.
2. Pratt, Lorien Y. Discriminability-Based Transfer between Neural Networks // NIPS 1992.
3. Vaswani A. et al. Attention Is All You Need // NIPS 2017.
4. Sutskever I., Vinyals O., Le Q. V. Sequence to Sequence Learning with Neural Networks // NIPS 2014.
5. Dong L., Xu S., Xu B. Speech-Transformer: A No-Recurrence Sequence-to-Sequence Model for Speech Recognition // 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2018.
6. Zhang Q. et al. Transformer Transducer: A Streamable Speech Recognition Model with Transformer Encoders and RNN-T Loss // NIPS 2020.
7. LeCun Y. et al. Backpropagation Applied to Handwritten Zip Code Recognition // Neural Computation. — 1989. — Vol. 1, no. 4. — Pp. 541–551.
8. Gulati A. et al. Conformer: Convolution-augmented Transformer for Speech Recognition // NIPS 2020.
9. Baevski A. et al. Wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations. ArXiv, 2020.
10. Fan A. et al. Addressing Some Limitations of Transformers with Feedback Memory // NIPS 2020.
11. Bhattamishra S., Ahuja K., Goyal N. On the Ability and Limitations of Transformers to Recognize Formal Languages // Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2020.
12. Ramsauer H. et al. Hopfield Networks is All You Need // NIPS 2020.
13. Тарков М. С. Нейрокомпьютерные системы. — М.: Интернет-Ун-т Информ. Технологий : Бином. Лаборатория знаний, 2006. — 142 с.
14. Hopfield J. J. Neural networks and physical systems with emergent collective computational abilities // Proceedings of the National Academy of Sciences. Proceedings of the National Academy of Sciences. — 1982. — Vol. 79, no. 8. — Pp. 2554–2558.
15. Krotov D., Hopfield J. J. Dense Associative Memory for Pattern Recognition // arXiv. 2016.
16. ml-jku/hopfield-layers: Hopfield Networks is All You Need. URL: <https://github.com/ml-jku/hopfield-layers> (дата обращения 24.03.2023).

17. sovaai/sova-dataset (sova RuYoutube). URL: <https://github.com/sovaai/sova-dataset> (дата обращения 14.02.2023).
18. voxforge.org. URL: <http://www.voxforge.org/ru> (дата обращения 14.02.2023).
19. *Karpov N., Denisenko A., Minkin F.* Golos: Russian Dataset for Speech Research // NIPS 2021.
20. dangrebenkin/audio_augmentator: A library to make augmented audio (16000 Hz, mono, WAV PCM) with noises corpora. URL: https://github.com/dangrebenkin/audio_augmentator (дата обращения 12.05.2023).
21. *Kürzinger L. et al.* CTC-Segmentation of Large Corpora for German End-to-End Speech Recognition // Speech and Computer. — 2020. — Pp. 267–278.
22. bond005/wav2vec2-large-ru-golos Hugging Face. URL: <https://huggingface.co/bond005/wav2vec2-large-ru-golos> (дата обращения 12.05.2023).

СРАВНЕНИЕ ФОРМАНТНЫХ ТРАЕКТОРИЙ ГЛАСНЫХ В РАЗЛИЧНЫХ ЯЗЫКАХ

Вера Вячеславовна Евдокимова

Мария Романовна Максимова

(Санкт-Петербургский государственный университет)

Введение. Для описания звука речи важны два параметра: движения органов артикуляции, необходимые для его произнесения, и участки речевого тракта, которые влияют на акустическую картину (спектр) звука. Спектр формируется при прохождении звука через надгортанные полости речевого тракта. Надгортанные полости являются резонаторами, усиливающими в спектре звука те частоты, которые совпадают с собственными частотами резонаторов. Эти усиленные частоты называются формантами. Собственная частота резонаторной полости зависит от ее формы, которая изменяется в зависимости от положения органов артикуляции, расположенных в этой полости [1].

Со второй половины XX века в фонетических исследованиях отмечается, что на качество гласного преимущественно влияют значения первой и второй формант. Значение первой форманты (F1) связано с движением языка по направлению к мягкому или твердому небу, то есть с подъемом гласного. В свою очередь, вторая форманта (F2) коррелирует с движением языка в полости рта по горизонтали, то есть с рядом гласного. Так как F1 и F2 имеют первостепенное значение для качества гласного, эти спектральные характеристики важны для распознавания гласных [3].

Акустическая фонетика изучает динамику формантных переходов гласных. В рамках одного из подходов к этому явлению, совокупность формантных переходов в речи одного диктора описывается как набор отдельных траекторий. Альтернативный подход предполагает рассмотрение формантных переходов, реализуемых одним диктором, в качестве элементов связной структуры. Преимущество этого подхода возможно подтвердить тем, что ключевую роль в восприятии звуков речи играют не статические характеристики, а динамика спектральных параметров во времени [2].

Формантные переходы имеют значение для разборчивости речи. Некорректные формантные переходы увеличивают вероятность ошибки при распознавании гласных [5].

Для распознавания особенно важно рассматривать формантные переходы внутри дифтонгов, а также внутри монофтонгов, которые имеют дифтонгоидную природу (таких, как английский напряженный гласный /i:/) [6].

Другим существенным параметром для распознавания гласных является угол наклона формантной траектории. Эта характеристика зависит от различий в положении речевых органов при произнесении соседних звуков и, следовательно, от значений формант этих звуков [6].

Целью данной работы является исследование вариативности формантных траекторий гласного в разных языках после губных, переднеязычных и заднеязычных согласных.

Материал и методика. Исследование проводилось на материале текстов, предложений и отдельных слов, прочитанных информантами женского пола на английском (британский вариант), французском и немецком языках. Аудиоматериал на английском языке был заимствован из корпуса LUCID (London UCL Clear speech in interaction) и из аудиоприложения к учебному пособию «MyGrammarLab: Advanced C1/C2». В материал на французском языке вошла запись, выполненная в СПбГУ в 2022 году с участием носителя французского языка. Также был использован аудиокурс к учебному пособию «EDITO C1 Méthode de français». Материал на немецком языке включал в себя рассказ и аудиокурс к учебному пособию «Wir-1».

В материал исследования были включены 12 английских, 10 французских и 16 немецких монофтонгов. Дифтонги и носовые гласные не были рассмотрены, так как дифтонги отсутствуют во французском языке, а в системах английских и немецких фонем не представлены носовые гласные.

Для построения графиков формантных траекторий были вычислены средние значения первой и второй формант в 9 точках гласного с помощью скрипта в программе Praat. Графики были построены с помощью программы Excel. Далее были вычислены производные полученных функций в каждой из 9 точек в целях сравнения крутизны формантных траекторий в английском, французском и немецком языках (рис. 1).

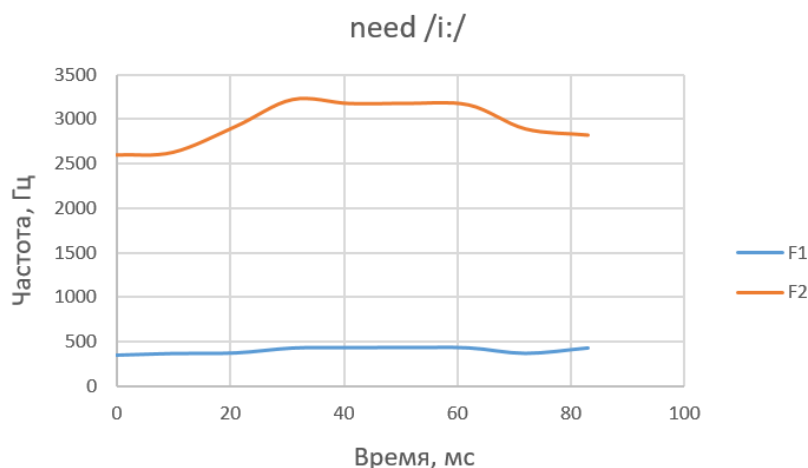


Рис. 1. График формантных траекторий F1 (нижняя кривая) и F2 (верхняя кривая) гласного /i/ в английском слове «need»

Для оценки вариативности изменения формантной структуры значения формант в Герцах были переведены в Барки. Это позволило понять, какие изменения формант были ощутимы для человеческого слуха.

Результаты по контурам формантных траекторий. Формантные траектории гласных больше всего различаются в контексте носовых согласных. Перед носовыми согласными наблюдаются как понижения, так и повышения формант. В немецких гласных наблюдаются более высокие значения F2 во всех контекстах. Однако значения F1 являются более высокими в английских гласных. На переходном участке во французских неогубленных гласных изменения формант были менее значительными, чем в других языках. Во французских огубленных гласных переходный участок был более резким, в том числе, после губных согласных. В немецких гласных, в целом, наблюдалось меньше изменений формант, чем в других языках во всех контекстах (рис. 2).

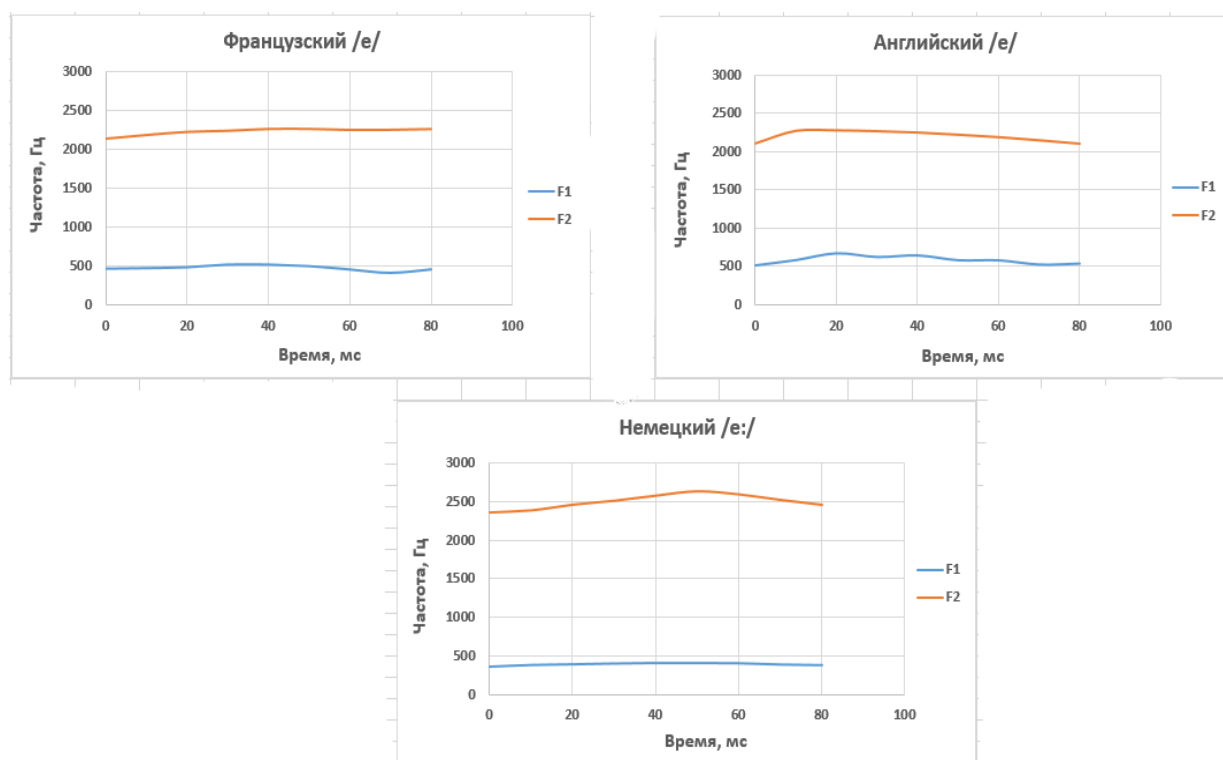


Рис. 2. Усредненные графики формантных переходов гласного /e/ (/e:/) после переднеязычных согласных во французском, английском и немецком языках F1 (нижняя кривая) и F2 (верхняя кривая)

Результаты по крутизне траекторий. В английском и немецком языках крутизна траекторий была наибольшей в гласных заднего ряда в любом контексте. Во французском языке крутизна траектории была более значительной в закрытых гласных переднего ряда. Наибольшая крутизна траектории в данном языке наблюдалась в центре или на втором переходном участке гласного. В английском и немецком языках, напротив, первый переходный участок характеризовался более значительной

крутизной траектории. В целом, крутизна переходов F2 была больше, чем крутизна траекторий F1 (рис. 3, 4).

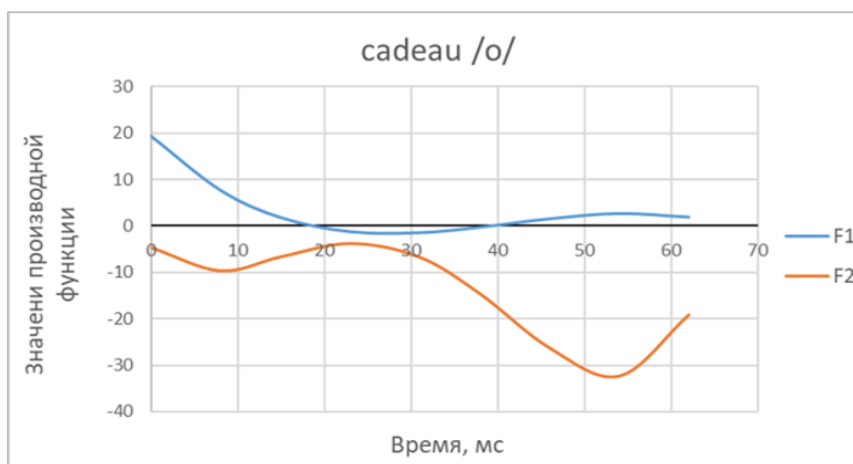


Рис. 3. График крутизны формантных переходов во французском гласном заднего ряда /o/. F1 (верхняя кривая) и F2 (нижняя кривая)

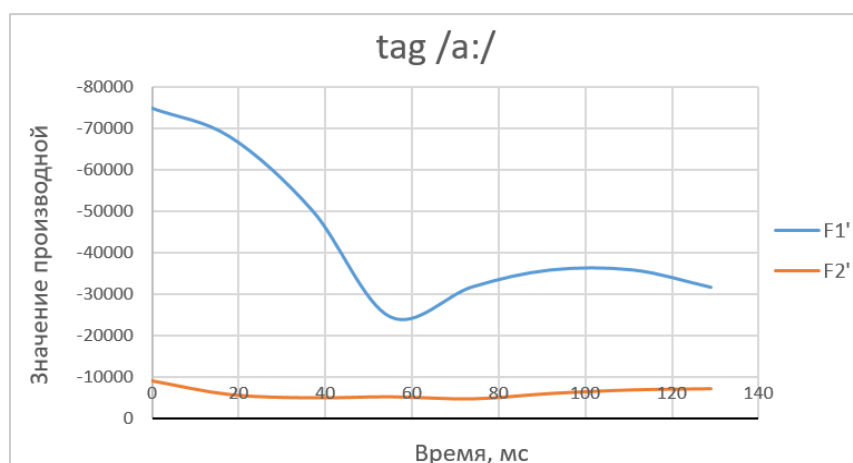


Рис. 4. График крутизны формантных переходов в немецком гласном заднего ряда /a:/. F1 (верхняя кривая) и F2 (нижняя кривая)

Результаты по вариативности формантной структуры. В большинстве случаев гласные в разных языках не отличались по вариативности формантной структуры. Наибольшая вариативность наблюдалась в F1 английских гласных. Вариативность в F1 была выше чем в F2. Значения F1 и F2 французских и немецких гласных чаще всего были выше, чем значения формант английских гласных.

В английских гласных заднего ряда наблюдалось меньше изменений в Барках, чем в гласных переднего ряда (табл. 1). Можно сказать, что с продвижением английских монофтонгов в сторону заднего ряда их частоты, воспринимаемые на слух, становятся более стабильными. Эта тенденция также прослеживалась на материале немецкого языка (табл. 2).

Таблица 1

**Усредненные значения F1 английского гласного /ɔ/ (/ɒ/)
в герцах и барках после заднеязычных согласных**

Английский /ɔ/	
Гц	Барки
566	5.0
705	6.0
594	5.2
617	5.4
606	5.3
589	5.2
519	4.7
480	4.4
595	5.3

Таблица 2

**Усредненные значения F1 немецкого гласного /ɔ/ (/ɒ/)
в герцах и барках после заднеязычных согласных**

Немецкий /ɔ/	
Гц	Барки
662	5.7
673	5.8
687	5.9
688	5.9
663	5.7
644	5.6
612	5.4
614	5.4
549	4.9

Таблица 3

**Усредненные значения F1 французского гласного /ɔ/ (/ɒ/)
в герцах и барках после заднеязычных согласных**

Французский /ɔ/	
Гц	Барки
643	5.6
561	5.0
629	5.5
661	5.7
621	5.4
647	5.6
683	5.9
587	5.2
629	5.5

Вариативность воспринимаемой частоты французских гласных, напротив, повышалась по мере их продвижения от переднего ряда к заднему (табл. 3). Однако во французском языке, в целом, реже наблюдались изменения по шкале Барков в другую по сравнению с английскими и немецкими монофтонгами.

Список использованной литературы

1. Основы общей фонетики [Текст]: учеб. пособие для студентов лингв. и филол. спец. / Л. В. Бондарко, Л. А. Вербицкая, М. В. Гордина. — 4-е изд., испр. — М.: Академия; СПб.: Филол. фак. СПбГУ, 2004.
2. *Broad D. J., Clermont F.* Target-locus scaling methods for modeling families of formant trajectories // *Journal of Phonetics*. — 2010, vol. 38. — Pp. 337–359.
3. *Fant G.* *The Acoustic Theory of Speech Production* (Mouton, The Hague), 1960.
4. *Lindblom B., Mauk C., Moon S.-J.* Dynamic specification in the production of speech and sign / eds P. Divenyi, S. Greenberg, G. Meyer // *Dynamics of speech production and perception* (NATO science series, series I: Life and behavioral sciences, Vol. 374). Amsterdam: IOS Press, 2006.
5. *Stachurski M., Summers R. J., Roberts B.* The verbal transformation effect and the perceptual organization of speech: influence of formant transitions and F0-contour continuity // *Hearing Research*. — 2015, vol. 323. — Pp. 22–31.
6. *Watson C. I., Harrington J.* Acoustic evidence for dynamic formant trajectories in Australian English vowels // *Journal of the Acoustical Society of America*. — 1999, vol. 106. — Pp. 458–468.

УНИВЕРСАЛЬНЫЕ И УНИКАЛЬНЫЕ ЧЕРТЫ В ИНТОНАЦИИ ФРАНЦУЗСКИХ ВОПРОСИТЕЛЬНЫХ ВЫСКАЗЫВАНИЙ

Ольга Вячеславовна Камрыш
(Санкт-Петербургский государственный университет)

Интонационная организация высказывания является неотъемлемой частью системы языка. Сопоставление описаний интонационных моделей, существующих в современной теоретической литературе, показывает, что общей чертой русского и французского языков является наличие синонимичности интонационных конструкций [3; 6; 11]. Универсалией следует также признать использование некоторых акустических средств, которые модифицируют интонационный контур, воспринимаемый как нейтральный, и позволяют передать эмоционально-оценочные значения [9]. Однако количество возможных вариантов мелодического оформления и их разнообразие, по-видимому, являются уникальными для каждого языка.

Интонационные конструкции русского языка и эмоционально-оценочные значения, которые могут быть переданы с помощью этих контуров, описаны в работах Е. А. Брызгуновой [1]. Применительно к французскому языку номенклатура интонационных моделей впервые была установлена П. Делаттром [10]. Система, разработанная П. Делаттром, была положена в основу большинства более поздних исследований французской интонации. Этот перечень включает в себя особый контур «импликация» (*implication*), имеющий восходяще-нисходящее движение тона и связанный с выражением дополнительных эмфатических значений. Все прочие интонационные конструкции привязываются к определенному синтаксическому типу высказывания. Следует отметить, что работы, посвященные интонационным характеристикам французских высказываний, нередко ограничиваются упрощенными схемами, которые показывают лишь общее направление мелодических изменений и не принимают во внимание вариативность интонационных контуров. При этом не ясно, какие интонационные конструкции французского языка могут иметь эмоционально-оценочные значения и стилистические варианты, подобно тому, как это происходит в русском языке. Таким образом, установление границ варьирования интонационных контуров во французском языке позволит выявить универсальные и специфичные черты интонационного оформления высказываний в русском и французском языках, что представляется существенным как для решения прикладных за-

дач автоматического синтеза речи, так и при определении лингводидактических основ преподавания иностранного языка. Кроме того, следует выделить особенности восприятия интонационного оформления высказываний разных синтаксических типов. В ходе специально организованных перцептивных экспериментов должны быть установлены интонационные черты тех высказываний, которые оцениваются носителями языка как «нейтральные». Важной задачей является также определение интонационных характеристик, которые могут быть соотнесены с дополнительными эмоциональными и модальными значениями.

Для решения поставленных задач совместно с обучающимися отделения «Общая и прикладная фонетика» СПбГУ была проведена серия экспериментально-фонетических исследований. Материалом послужили аудиодорожки к современным французским фильмам, подкасты, находящиеся в свободном доступе, а также аудиозаписи к аутентичным методическим пособиям уровня не ниже В2 (в соответствии с шкалой общеевропейских компетенций владения иностранным языком). Общая длительность обследованного материала составила более 55 часов звучания. Всего было рассмотрено 559 вопросительных высказываний. Анализу подлежали французские вопросительные предложения разной синтаксической организации: 1) частные вопросы с вопросительным словом в начале предложения и оборотом *est-ce que*; 2) частные вопросы с прямым порядком следования подлежащего и сказуемого и вопросительным словом в конце высказывания; 3) частные вопросы, имеющие прямой порядок слов и вопросительное слово в начале предложения.

Результаты проведенного экспериментального исследования дают основание говорить о том, что традиционно рекомендуемая модель с нисходящим движением тона от начала к концу высказывания [4; 10; 11] не является единственно возможной в современном французском языке. При анализе частных вопросов с разной синтаксической организацией было показано, что существуют также иные способы мелодического оформления, в том числе с использованием ровного и восходящего тона [5]. Кроме того, более детальное описание необходимо сделать и в случае реализации нисходящих мелодических контуров. Так, для высказываний с вопросительным словом в начале предложения и прямым порядком следования подлежащего и сказуемого вариант мелодического оформления с нисходящим движением тона был зафиксирован в 82 % случаев [8]. Однако, вариативность мелодических изменений внутри группы нисходящих контуров оказалась более значительной, чем следует из описаний в методической и теоретической литературе. Изучение интонационного оформления вопросительных предложений данного синтаксического типа показывает, что во внимание, по-видимому, следует принимать не только направление движения тона от начала к концу

высказывания, как это сделано в классификации П. Делаттра, но и интонационное оформление отдельных участков высказывания. Представляется возможным выделить информативно значимые отрезки интонационной кривой. К ним следует отнести участок вопросительного слова, а также финальный (финальные) слог (слоги) предложения. На рисунках 1–3 показаны мелодические варианты вопросительных высказываний, выявленные в результате проведенного акустического анализа.

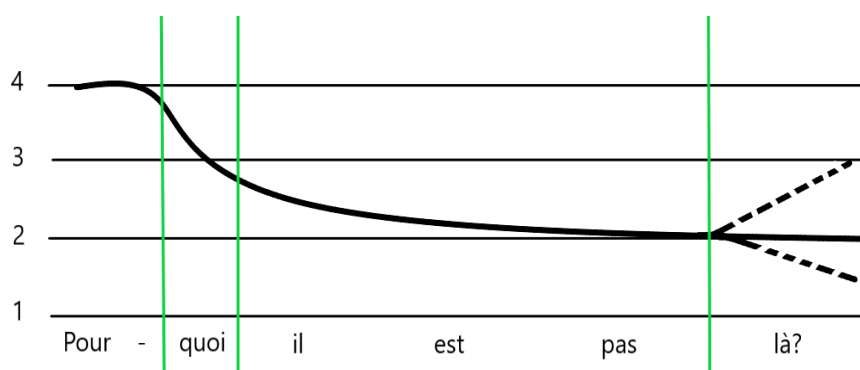


Рис. 1. Реализация частного вопроса с падением на вопросительном слове

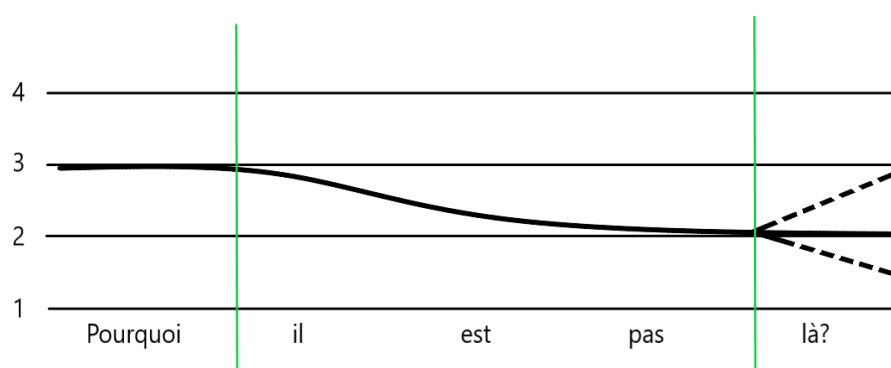


Рис. 2. Реализация частного вопроса с ровным тоном на вопросительном слове

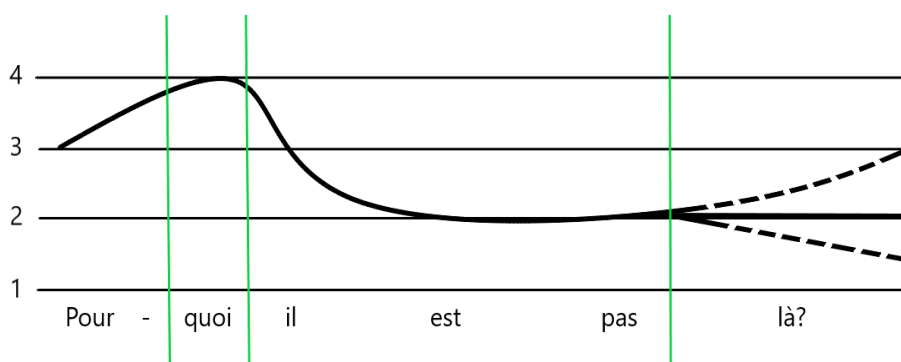


Рис. 3. Реализация частного вопроса с подъемом тона на вопросительном слове

Учет направления и характера движения частоты основного тона, прежде всего, на этих отрезках интонационного контура позволяет выявить все многообразие интонационных вариантов, которое существует

в современной речевой практике носителей французского языка (рисунки 1–3). Кроме того, может быть поставлен вопрос о связи этих мелодических изменений с восприятием дополнительных модальных значений и различных эмоциональных оттенков высказывания.

Определение фонологически существенных интонационных различий следует проводить лишь при обращении к восприятию носителей языка. Для выявления эмфатических или модальных значений, приписываемых тому или иному высказыванию, была организована серия перцептивных экспериментов. На основании результатов перцептивного анализа были получены следующие предварительные выводы. Во-первых, можно констатировать, что оценка степени заинтересованности говорящего в предмете разговора связана с характером мелодического оформления последнего ударного слога или двух последних слогов высказывания. Интенсивный подъем на ударном слоге соотносился аудитором с выражением «приветливости» и «заинтересованности». При изменении интонации ударного слога ресинтезированных вопросов с резко восходящей на ровную или нисходящую участники перцептивного эксперимента отмечали «безразличие» в голосе диктора, а обстановка общения характеризовалась в целом как «холодная, недружелюбная». Во-вторых, была установлена корреляция между формой мелодической кривой и восприятием тех или иных эмоций. Прежде всего, необходимо отметить, что появление дополнительных оценок в ответах аудиторов может быть связано с изрезанностью интонационного контура. В подобных высказываниях участники перцептивного эксперимента указывали наличие негативных эмоций: «раздражение», «требовательность», «испуг». В-третьих, полученные данные позволяют говорить о влиянии других акустических параметров на оценку высказывания как эмфатического. Прежде всего, результаты перцептивного анализа показывают, что существенным для восприятия фактором является расширение или, напротив, сужение диапазона варьирования значений частоты основного тона внутри высказывания в целом или же на отдельных его участках. Например, высказывания с вопросительным словом в начале предложения и прямым порядком слов были оценены как произнесенные с эмоцией «удивления» при увеличении диапазона мелодических изменений на вопросительном слове или в финале фразы. Оpozнание «удивления» при этом было зафиксировано как для фраз, имеющих нисходящее движение тона на вопросительном слове, так и с подъемом на этом участке высказывания. Таким образом, решающее значение для восприятия дополнительного значения «удивление», по-видимому, имеет именно величина диапазона значений частоты основного тона, а не направление движения мелодики. Очевидно, что в этом случае данные аудиторского экспери-

мента подтверждают, высказанные ранее положения на материале других языков, в том числе русского, что может свидетельствовать об универсальности использования мелодических средств для передачи эмфатических значений. Так, связь эмоциональной оценки с изрезанностью и сложностью мелодического контура была отмечена для русского языка в работе Н. Д. Светозаровой [7]. Изменение диапазона мелодических изменений и регистровых значений по сравнению с нейтральной реализацией также неоднократно указывались как отличительная черта эмоционально-окрашенных высказываний в разных языках [9; 12].

Напротив, за сходным в мелодическом оформлении восходяще-нисходящим контуром в речи носителей русского и французского языков закреплены разные значения. Наличие восходяще-нисходящего движения тона при оформлении вопросительного высказывания и незавершенности (ИК-3) традиционно считается характерной чертой интонационной системы русского языка [2]. Полученные экспериментальные данные свидетельствуют о том, что во французском языке подобный контур используется для передачи дополнительных эмоциональных значений. Так, в набор оттенков, отмеченных респондентами, вошли «любопытство» и «удивление»; тон говорящего получил такие положительные эмоциональные оценки как «счастливый», «дружелюбный», «восхищенный», «уверенный в себе». Таким образом, колоколообразная форма мелодической кривой на интонационном центре, которая может быть соотнесена с ИК-3 в классификации Е. А. Брызгуновой [1], по-видимому, является одним из способов выражения некоторых эмоциональных значений во французском языке в отличие от русского языка, где восходяще-нисходящее движение тона свойственно эмоционально-нейтральным высказываниям.

Список использованной литературы

1. Брызгунова Е. А. Звуки и интонация русской речи. — М.: Русский язык, 1977. — 281 с.
2. Вольская Н. Б. Интонация и языковой контакт: прагматический аспект внутри- и межъязыковой интерференции // XLII Международная филологическая конференция. Избранные труды. Санкт-Петербургский государственный университет 2014. — С. 85–92.
3. Ефремова М. Ю. Синонимия интонационных конструкций в повествовательных и вопросительных предложениях // Вестник Московского государственного лингвистического университета. Гуманитарные науки. — 2021, вып. 11. — С. 84–97.
4. Гордина М. В. Фонетика французского языка. — СПб.: Изд-во С.-Петербург. ун-та, 1997. — 304 с.
5. Камрыш О. В. К вопросу о вариативности мелодических контуров в современном французском языке // Обучение иностранному языку студентов высших и средних

- образовательных учреждений на современном этапе: Материалы IX Всероссийской национальной научно-методической видеоконференции (с международным участием). — Благовещенск: Изд-во Амурского гос. ун-та, 2021. — С. 91–97.
6. *Рубер М. М.* Интонационные конструкции в русском и французском языках // Русский язык за рубежом. — 1984, № 4. — С. 78–82.
 7. *Светозарова Н. Д.* Интонационная система русского языка. — Л.: Изд-во Ленингр. ун-та, 1982. — 175 с.
 8. *Северинова А. А.* Интонационная вариативность французских частных вопросов и особенности их восприятия. Выпускная квалификационная работа. — СПб., 2021. — 127 с.
 9. *Юрова И. В.* Интонационные универсалии и восприятие эмоциональных и неэмоциональных интонаций (экспериментально-фонетическое исследование на материале французского языка). — СПб., 1984. — 352 с.
 10. *Delattre P.* Les dix intonations de base de français // *The French Review*. — 1966. — Vol. 40, no. 1. — Pp. 1–14.
 11. *Léon P.* Phonétisme et prononciations du français: avec des travaux pratiques d'application et leurs corrigés. — Paris: Nathan Université, 1996. — 192 p.
 12. *Vaissière J.* Language independent prosodic features // *Prosody: Models and Measurements*. — Springer Verlag, 1983. — Pp. 53–65.

СОПОСТАВИТЕЛЬНЫЙ АНАЛИЗ ДЫХАТЕЛЬНОЙ АКТИВНОСТИ ДИКТОРА В СОСТОЯНИИ ПОКОЯ И ПРИ ЧТЕНИИ ПРОЗАИЧЕСКОГО И ПОЭТИЧЕСКОГО ТЕКСТОВ

Галина Евгеньевна Кедрова

Мария Владимировна Волкова

Николай Викторович Анисимов

(Московский государственный университет им. М. В. Ломоносова)

В статье представлены результаты сопоставительного изучения дыхательной практики дикторов в спокойном состоянии (без выполнения каких-либо речевых заданий) и при чтении разных типов текстов (прозаического и стихотворного) по МРТ-изображениям лёгких, сделанным в режиме реального времени (онлайн). В предыдущих наших публикациях было высказано предположение, что речевое дыхание требует существенного увеличения объёма воздуха в лёгких по сравнению с работой лёгких в спокойном состоянии [1]. Также существует гипотеза, что при выполнении разных речевых задач объём лёгких меняется менее значительно, так как говорящий стремится поддерживать некоторый объём воздуха, критически важный для поддержания речевой активности [2]. Для экспериментальной проверки этих предположений была проведена специальная обработка онлайн-МРТ-изображений с целью получения численных значений размера наиболее активных участков лёгких в молчании при прочтении поэтического текста и при чтении короткого рассказа. Полученные данные были подвергнуты статическому анализу, а для оценки значимости/незначимости различий в полученных выборках использовался критерий Манна-Уитни (U-критерий) [2].

Лингвистическим материалом для онлайн-МРТ-визуализации работы лёгких при чтении текста послужили текстовые фрагменты, приведенные в книге Р. И. Аванесова «Русское литературное произношение» — два стихотворных текста («Парус» М. Ю. Лермонтова и «Словарь» С. Маршака), а также один прозаический — «Говорящий грач» М. Пришвина [3]. Инструментальное исследование физиологии дыхательного процесса в режиме онлайн в покое (молчании) и при чтении экспериментальных текстов было проведено в ЛМТС МГУ на магнитно-резонансном томографе Tomikon S50 «Bruker» (0.5 Тл). Параметры МРТ-сканирования: применялся метод градиентного эхо (GRE) с разрешением 6.7×6.7 мм² в коронарной проекции без срезовой селекции в режиме эволюции — 200 кадров. При параметрах TR/TE = 5.9/2.2 мс время сканирования зоны интереса (кадра) составило 0.213 сек, задержка между кад-

рами — 0.2 с. В результате период следования кадров составил 0.413 с, общее время их сканирования в каждой из сессии запуска аппарата — 82.6 с (1 мин. 23 с). В течение этого времени испытуемый находился в спокойном состоянии (молчание) или читал заготовленный текст.

Всего в МРТ-экспериментах приняли участие 5 дикторов (2 мужчин и 3 женщины). Подробная методика проведения МРТ-экспериментов и образец инструкции для дикторов в каждой из экспериментальных сессий, а также технология первичной обработки (идентификация дыхательных периодов) и темпоральной привязки полученных данных представлены в работе [4]. По техническим условиям МРТ-сканирования дикторы должны были читать экспериментальные тексты, лежа на спине. Для наиболее детального отображения активности лёгких испытуемых просили поднять и зафиксировать обе руки над головой. Дикторы могли предварительно ознакомиться с напечатанным на бумаге текстом, однако ни один из них не получал от экспериментатора никаких указаний по выбору какого-либо особого декламационного типа чтения. Текст для чтения располагался непосредственно перед глазами диктора на удобном для каждого говорящего расстоянии. Все тексты (как поэтический, так и прозаический) были прочитаны дикторами в ходе двух сеансов МРТ-съёмки для каждого вида текста, между сеансами делались короткие паузы для отдыха диктора. Для МРТ-съёмки дыхания в покое был проведен один сеанс. Таким образом, в процессе эксперимента по каждому из дикторов было проведено 5 серий запусков последовательности МРТ-сканирования, которые позволили получить покадровые изображения лёгких в покое (молчание), при чтении прозаического текста (два сеанса) и при чтении стихотворного текста (два сеанса). Последовательность текстов для прочтения (сначала прозаический текст, потом стихотворный/сначала стихотворный текст потом прозаический) менялась в каждом сеансе сканирования для каждого из диктора с целью проверки устойчивости индивидуальных моделей чтения каждого типа текста. В итоге для каждого вида речевой деятельности по каждому диктору было получено по 400 МРТ-изображений работы лёгких + 200 МРТ-изображений спокойного дыхания (молчание).

Проведенная обработка данных МРТ не выявила существенной междикторской вариабельности. Поэтому в настоящей работе приведены эти результаты для одного из них — диктора-женщины. На рис. 1 представлены МРТ-кадры, на которых зафиксированы фазы максимального и минимального объёма воздуха в лёгких при дыхании в молчании и в ходе выполнения речевых задач диктором 1.

Для первичной визуальной оценки МРТ-изображений дыхательного процесса в покое и при выполнении речевых задач проводилась специальная обработка полученных кадров в графическом пакете Photoshop

(Adobe), которая заключалась в подборе оптимального режима автоматической трассировки изменения интенсивности сигнала в зоне на границе обоих лёгких. Результат такой обработки МРТ-кадров представлен на рис. 2. На следующем этапе полученные контуры, отмасштабированные к размеру 400×400 пикселей, обрабатывалась в стандартном графическом пакете ImageJ для подсчёта суммы площадей активных участков левого и правого лёгких с целью дальнейшей статистической обработки.

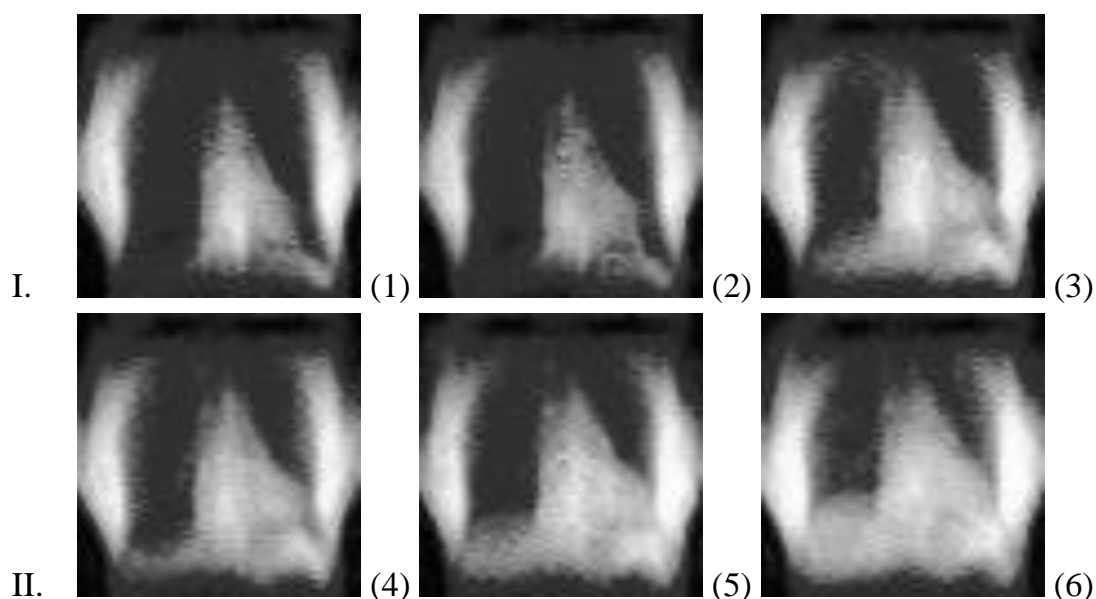


Рис. 1. I. Фаза максимального вдоха при чтении поэтического текста (1); при чтении прозаического текста (2); в молчании (3). II. Фаза максимального выдоха при чтении поэтического текста (4); при чтении прозаического текста (5); в молчании (6)

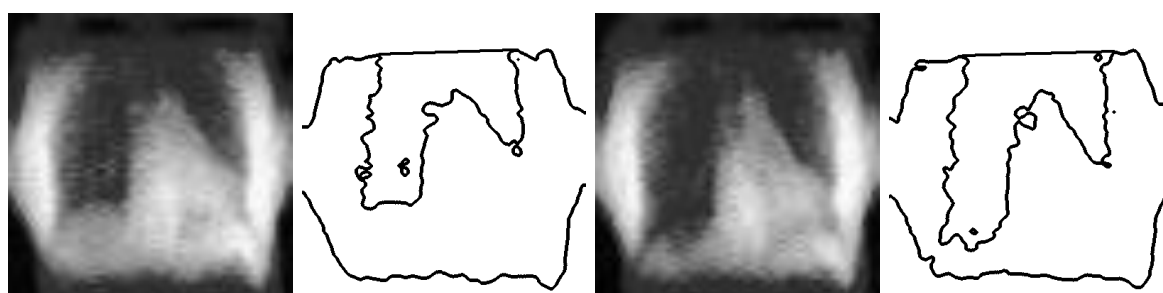


Рис. 2. МРТ-кадры до и после обработки в графическом пакете Photoshop (Adobe)

В результате проделанных операций был получен массив данных, отображающий изменение общей площади активных участков лёгких как в процессе дыхания в молчании, так и при двукратном прочтении обоих видов экспериментальных текстов (по 200 МРТ-изображениям в каждой серии). Для каждой серии были подсчитаны медианное значение общего объёма лёгких в каждом из дыхательных режимов, а также минимальное и максимальное значения. Графики, представленные на 30

рис. 3, указывают на существенное увеличение объёма воздуха в лёгких при речевом дыхании по сравнению с работой лёгких при молчании (в среднем в 1,6 раза). Из графика также следует, что объём лёгких при выполнении разных речевых задач меняется незначительно, что подтверждает вторую из вышеуказанных гипотез.

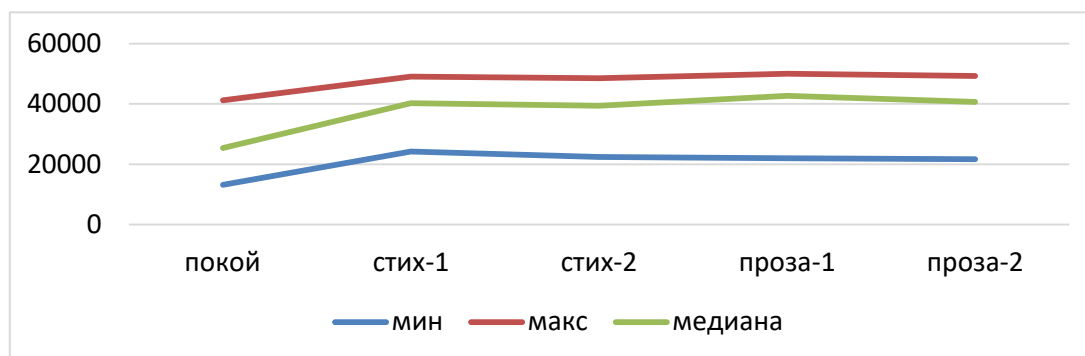


Рис. 3. Медианные, минимальные и максимальные значения объема лёгких (в пикселах) при дыхании в покое, в 1-ой сессии чтения поэтического текста (стих-1), во 2-ой сессии чтения поэтического текста (стих-2), в 1-ой сессии чтения прозаического текста (проза-1), во 2-ой сессии чтения прозаического текста (проза-2). На оси ординат — значения общей площади активных участков обоих лёгких в пикселах

Для сравнения интегрального представления данных, полученных в результате подсчета площади лёгких по каждому МРТ-контур, использовалась диаграмма Дж. Тьюки (рис. 4), что позволяет сопоставить в удобной наглядной форме несколько выборок, представив по каждой выборке медиану, нижний и верхний квартили, а также минимальное и максимальное значения и наличие выбросов.

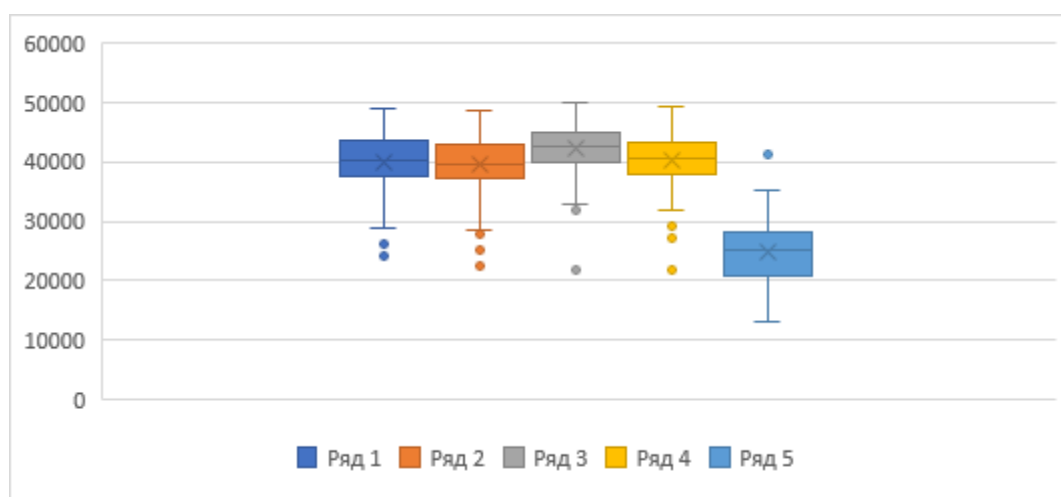


Рис. 4. Ряд 1 — 1-ая сессия чтения поэтического текста; ряд 2 — 2-ая сессия чтения поэтического текста; ряд 3 — 1-ая сессия чтения прозаического текста; ряд 4 — 2-ая сессия чтения прозаического текста; ряд 5 — дыхание в молчании. На оси ординат — значения общей площади активных участков обоих лёгких в пикселах

Представленные на графиках результаты подтверждают гипотезу о значительном системном отличии дыхательной стратегии, используемой говорящим при производстве речи, от дыхания в молчании. Для статистической проверки этой гипотезы мы использовали непараметрический критерий Манна-Уитни (U-критерий), применяемый в психологии для сравнения двух независимых выборок по уровню какого-либо количественно измеренного признака (<https://www.psychol-ok.ru/statistics/mann-whitney/>). [6] Результаты подсчета U-критерия для чтения поэтического текста, чтения прозаического текста и дыхания в молчании представлены в таблице 1. Для сравнения в таблице 1 также приведены значения U-критерия для двух серий прочтения поэтического текста.

Таблица 1

Значения U-критерия для разных типов текста в сопоставлении с дыханием при молчании

Выборка	Полученное эмпирическое значение U-критерия	Вывод о различии
Поэтический текст (1 чтение) / дыхание в молчании	$U_{Эмп} = 71$	находится в зоне значимости
Поэтический текст (2 чтение) / дыхание в молчании	$U_{Эмп} = 71$	находится в зоне значимости
Прозаический текст (1 чтение) / дыхание в молчании	$U_{Эмп} = 42$	находится в зоне значимости
Прозаический текст (2 чтение) / дыхание в молчании	$U_{Эмп} = 82$	находится в зоне значимости
Поэтический текст (1 чтение) / поэтический текст (2 чтение)	$U_{Эмп} = 1534$	находится в зоне незначимости

Таким образом, наше экспериментальное исследование дыхательной активности диктора при чтении прозаического и поэтического текстов, выполненное с использованием онлайн-ового МРТ-сканирования лёгких, показало системное отличие речевого дыхания от дыхания в молчании. Важно, что данные онлайн-овой МРТ-визуализации дыхательных процессов при производстве речи позволяют количественно оценить специфику речевого дыхания как значимого физиологического фактора в рамках процесса производства речи.

Список использованной литературы

1. *Кедрова Г. Е., Анисимов Н. В., Волкова М. В.* Новые данные о речевом дыхании по результатам онлайн-магнитно-резонансной томографии лёгких при чтении поэтического текста // ...Вперёд и вверх по лестнице звучащей. Сборник статей к 80-летию Ольги Фёдоровны Кривновой / под ред. Л. М. Захаров, И. М. Кобозева, А. Э. Костюк и др. — М.: Москва, 2023. — С. 260–270.
2. *Зиндер Л. Р.* Общая фонетика. — М.: Высшая школа, 1972. — 312 с.
3. *Норина В. Н.* Критерий Манна-Уитни в психологии // Образовательный портал «Справочник». URL: https://spravochnick.ru/psihologiya/kriteriy_manna-uitni_v_psihologii/ (дата обращения: 07.05.2023).
4. *Аванесов Р. И.* Русское литературное произношение. — М.: Просвещение, 1968. — 286 с.
5. *Кедрова Г. Е., Анисимов Н. В.* Новые данные о речевом дыхании по результатам онлайн-магнитно-резонансной томографии легких // Сборник трудов XXXIV сессии Российского акустического общества. — М.: ГЕОС, 2022. — С. 96–102.
6. *Сидоренко Е. В.* Методы математической обработки в психологии. СПб.: ООО Речь, 2007. URL: <https://www.psychol-ok.ru/statistics/mann-whitney/> (дата обращения: 07.05.2023).

ИНТОНАЦИЯ ЗАВЕРШЕННОСТИ И НЕЗАВЕРШЕННОСТИ В РУССКОЙ РЕЧИ БРАЗИЛЬЦЕВ

Ульяна Евгеньевна Кочеткова

Татьяна Васильевна Качковская

Арина Антоновна Соболева

(Санкт-Петербургский государственный университет)

Важным аспектом в изучении иностранного языка является освоение его интонационной системы, поскольку интонация отвечает не только за передачу коммуникативного типа высказывания, но и за выражение оттенков смысла и эмоциональной окраски. Передать свои мысли и переживания на иностранном языке с той же точностью, что и на родном языке, можно лишь владея навыками правильного интонационного оформления.

При обучении интонации принято опираться на закрытый список интонационных конструкций, для каждой из которых характерен свой мелодический рисунок и определенная функциональная нагрузка. В различных языках набор таких конструкций, как правило, различается, однако отдельные интонационные конструкции могут быть схожими. Но даже при внешнем сходстве мелодического оформления могут присутствовать функциональные различия, с чем и связаны основные сложности, возникающие при обучении интонации неродного языка. Возникающая как в первом, так и во втором случае интонационная интерференция остается мало изученной, в отличие от интерференции в области лексики, грамматики и сегментной фонетики. Несмотря на большое число изучающих русский язык в Бразилии, до сих пор отсутствуют сведения об особенностях интонационного оформления русской речи бразильцами. В связи с этим целью настоящего исследования является анализ интонации завершенности и незавершенности в интерферированной речи бразильцев, поскольку в данных коммуникативных типах можно предположить возникновение отрицательной интерференции на супraseгментном уровне.

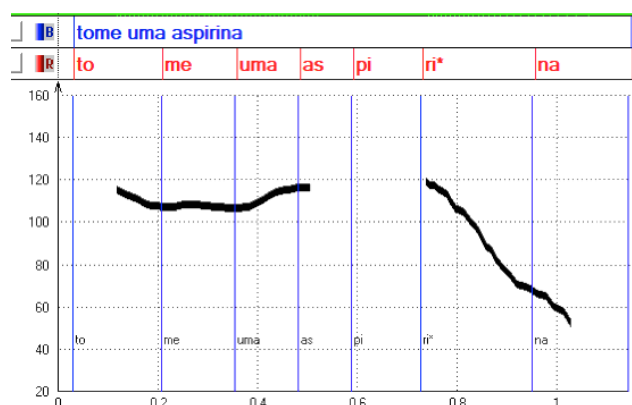
В качестве материала были отобраны 20 фраз из корпуса бразильской разговорной речи, разработанного на кафедре фонетики и методики преподавания иностранных языков СПбГУ [2]. Фразы были начитаны шестью бразильцами, владеющими русским языком на уровне А2-В2, но не прошедшими специального обучения интонации. Помимо фраз на русском языке были проанализированы фразы с аналогичным значением (и лексическим составом) на бразильском португальском. Кроме того,

было проведено сопоставление с фразами на русском, прочитанными двумя носителями русского языка (мужчиной и женщиной). Это позволило провести сравнение интонационного оформления идентичных по содержанию и коммуникативному типу фраз на русском и португальском языках, прочитанных одними и теми же дикторами — носителями бразильского португальского языка, а также сопоставить нормативное и интерферированное оформление конечных и неконечных синтагм повествовательных предложений.

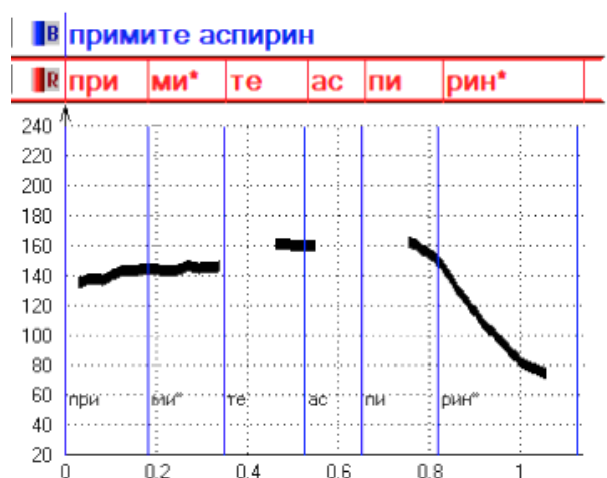
На первоначальном этапе была произведена аннотация и сегментация всего материала. Фонетическая аннотация включала в себя определение границ слогов и орфографическую расшифровку. Во всем материале была определена частота основного тона — метки периодов основного тона расставлялись вручную в программе Wave Assistant. Рассматривалось движение тона как в интонационном центре, так и в предупредном и ударном слогах. Далее на основании акустического анализа была произведена транскрипция в системе P-ToVi [7]. Для обозначения движения тона в зоне интонационного центра использовался комплекс из двух буквенных символов (L — низкий тон, H — высокий тон), в некоторых случаях добавлялись диакритики (! — более низкий уровень тона, ¡ — более высокий уровень тона). Буквенные символы объединялись знаком плюс (+). На символе, непосредственно относящемся к интонационному центру, ставился знак звездочка (*). Уровень тона предупредного и ударного слога обозначались аналогичными символами (L, H и диакритики) справа и/или слева от записи тона ударного слога. Граница интонационной единицы (синтагмы) обозначалась с помощью знака процент (%) (подробное описание системы P-ToVi дано в [6]).

Интонация завершенности. Повествовательные предложения, не имеющие особой эмоциональной окраски, оформляются в целом нисходящим движением тона в обоих языках. Однако в русском языке такая конструкция L* L% соответствует ИК-1 в системе Е. А. Брызгуновой [1] или интонационным моделям 01, 01a, 01b в системе Н. Б. Вольской [3]. Данная ИК характеризуется более или менее заметным падением на ударном слоге без подъема на предупредном. Это отличает данную конструкцию от ИК-2 (моделей 04, 02 в системе Н. Б. Вольской), которую можно представить как H+L* L% в терминах P-ToVi, и для которой свойственно падение с более высокого уровня. Согласно описанию Ж. А. Морайша [5], в бразильском португальском интонационный контур, характерный для нейтрального повествовательного предложения, начинается, наоборот, с высокого уровня H+L* L%, что объединяет его с русской ИК-2 (рис. 1a). Соответственно, можно предположить возникновение интерференции, выражающейся в замене одной модели на другую, что в русском языке будет приводить к появлению эмфазы и эмоциональной

окраски. Как и следовало ожидать, анализ русской речи бразильцев показал, что большая часть высказываний, прочитанных бразильцами на русском языке, была оформлена мелодическим контуром $H+L^* L\%$ (рис. 1б). Можно заметить, что при прочтении данной фразы на русском языке диктор воспроизводит такой же мелодический рисунок, что и в родном португальском, но смещает его в более высокий регистр (рис. 1а, б).



а)



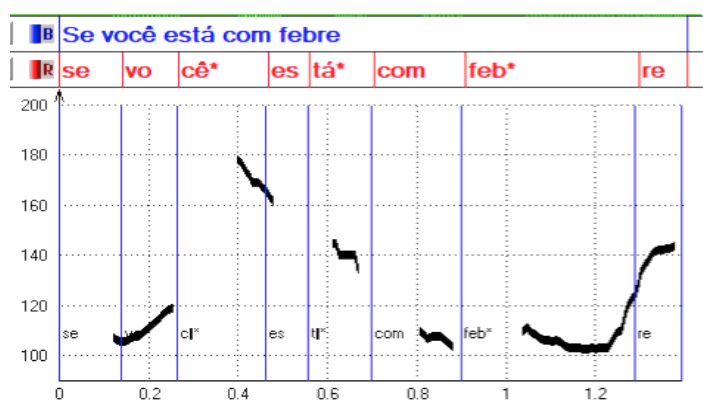
б)

Рис. 1а, б. Пример оформления конечной синтагмы «примите аспирин» (“tome uma aspirina”) бразильским диктором в португальском (а) и в русском (б) языках

Интонация незавершенности. В классификации, приведенной Л. К. Кальяри [3], незавершенность выражается в бразильском португальском за счет падения тона на интонационном центре с дальнейшим подъемом $H L+N^* H\%$ и соответствует русской ИК-4 ($L^*+H H\%$) в аналогичном значении (рис. 2а). Однако в русском языке интонация незавершенности чаще выражается с помощью ИК-3 ($L+N^*+L L\%$), а также ИК-6 ($L+N^* H\%$), в то время как для бразильского португальского другие

варианты мелодического контура не описаны. Следовательно, ожидаемым вариантом оформления неконечных синтагм в русской речи бразильцев может быть именно ИК-4.

В ходе анализа мелодических контуров в интерферированной речи были получены неожиданные результаты. Оказалось, что при перечислении и в других случаях выражения незавершенности использовался не только контур наподобие ИК-4 ($L^*+N\ N\%$), но и контур, который можно сравнить с ИК-3 ($L+N^*+L\ L\%$) (Рис. 2б), хотя такой контур в бразильском португальском, согласно известным классификациям, соответствует интонации общего вопроса [5].



а)



б)

Рис. 2. Пример оформления неконечной синтагмы «Если у вас высокая температура» (“Se você está com febre”) бразильским диктором в португальском (а) и русском (б) языках

Важной особенностью интонационного оформления неконечных синтагм в португальском языке является наличие слов с дополнительной просодической выделенностью, на ударных слогах которых наблюдаются заметные перепады тона (рис. 2а). При этом данный мелодический рисунок характерен для эмоционально-нейтральных высказываний. Интересно, что при реализации подобных синтагм на русском языке дик-

торы-бразильцы реализовывали еще большее количество таких слогов (рис. 2б).

Если же обратиться к типичному интонационному оформлению незавершенности в речи носителей русского языка, можно заметить резкий скачок тона на лишь на интонационном центре с дальнейшим спадом на заударном слоге, иногда почти доходящим до уровня тона предупредного слога (рис. 3).



Рис. 3. Пример оформления неконечной синтагмы «Если у вас высокая температура» носителем русского языка

В ходе анализа мелодического оформления неконечных синтагм при перечислении была замечена междикторская и внутридикторская вариативность, что не позволяет говорить о строгой корреляции между стратегиями интонационного оформления данного типа незавершенности в бразильском португальском и русском языках.

Подводя итоги настоящего исследования, можно сделать вывод о том, что интонация завершенности выражается в интерферированной речи бразильцев так же, как и в португальском, и имеет схожие черты не с ИК-1, а с ИК-2, из-за чего можно ожидать неправильную интерпретацию нейтральных повествовательных предложений с возможными дополнительными эмоциональными коннотациями. В некоторых случаях — при смещенном тайминге и реализации позднего подъема — возможно даже пересечение с ИК-3, приводящее к ошибочному восприятию коммуникативного типа.

Анализ интонации незавершенности в португальском и русском языках в речи бразильцев показал, что неконечные синтагмы оформляются с помощью интонационных конструкций, похожих на русские. Отличием при этом является наличие большего числа тональных пиков в синтагмах, состоящих из нескольких слов, что может затруднить правильное определение интонационного центра носителями русского языка. Во фразах с перечислением интонационная интерференция

оказалась наименее выраженной, что говорит о меньшей значимости изучения этой конструкции при обучении бразильцев русскому языку.

Список использованной литературы

1. *Брызгунова Е. А.* Звуки и интонация русской речи: [Для иностранцев, изуч. рус. яз.]. — 3-е изд., перераб. — М.: Рус. яз., 1977.
2. BraPoRus, spoken corpus of heritage Russian in Brazil: protocol of data collection / Smirnova Henriques A. [et al.] // *Cadernos de Linguística*. — 2022. — Vol. 3, no. 1. — P. 629.
3. *Вольская Н. Б., Скрелин П. А.* Система интонационных моделей для автоматической интерпретации интонационного оформления высказывания: функциональные и перцептивные характеристики // *АРЗ-2009*. — С. 28–40.
4. *Cagliari L.C.* Elementos de fonetica do portugues brasileiro. 1981. 192f. Tese (livre-docencia) — Universidade Estadual de Campinas, Instituto de Estudos da Linguagem, Campinas, SP. URL: <https://hdl.handle.net/20.500.12733/1577156> (дата обращения: 11 abr. 2023).
5. *Moraes J. A. d.* The pitch accents in Brazilian Portuguese: analysis by synthesis // *Proc. Speech Prosody 2008*. — Pp. 389–397.
6. *Соболева А. А.* Сопоставление интонационных систем бразильского португальского и русского языков. Выпускная квалификационная работа на соискание степени магистра по направлению “Лингвистика”, 2023.
7. <https://labfon.letras.ulisboa.pt/InAPoP/P-ToBI/>

АНАЛИЗ ЭМОЦИОНАЛЬНОЙ РЕЧИ ДЕТЕЙ¹

Елена Евгеньевна Ляско

Ольга Владимировна Фролова

Александр Сергеевич Николаев

(Санкт-Петербургский государственный университет)

В работе рассматриваются подходы для получения эмоциональной речи детей, данные перцептивных экспериментов по распознаванию эмоциональных состояний детей по характеристикам их голоса и речи, акустические характеристики эмоциональной речи.

Для получения эмоциональной речи детей разработан комплексный подход «Методика оценки эмоционального развития детей» (Child Emotion Development Method — CEDM), включающий тесты, задания и игровые ситуации, направленные на отражение детьми в голосе, речи и мимике разных эмоциональных состояний (ОЭ) и распознавание эмоциональных состояний других [1]. Дополнительно использованы аудио и видеозаписи, содержащиеся в базе данных «AD_CHILD.RU», созданной для изучения речевого онтогенеза и дизонтогенеза [2].

Речевой материал детей аннотирован двумя экспертами (коэффициент каппа — 1,0). В зависимости от задачи исследования аннотирование материала проводили по 3 состояниям «комфорт — нейтральное (спокойное) — дискомфорт» и по 4 состояниям «радость — нейтральное — печаль — гнев».

Для анализа эмоциональной речи детей использовали метод перцептивного эксперимента и спектрографический анализ.

Серии перцептивных экспериментов проведены с целью определения возможности распознавания разными группами взрослых, прослушивающих речевой материал детей (аудиторов), эмоциональных состояний детей. В кросс-культурном исследовании в качестве аудиторов выступали носители русского и тамильского языков. Кросс-лингвистические исследования являются наиболее сложными при изучении отражения эмоций в характеристиках голоса. Анализ акустических характеристик эмоциональной речи детей осуществлен в звуковом редакторе “Cool Edit Pro”. Для слов и высказываний детей определяли длительность, значения частоты основного тона (ЧОТ, F₀), максимальные значения ЧОТ (ЧОТ max, F₀max) и минимальные значения ЧОТ (ЧОТ min, F₀min), считали значения диапазона ЧОТ — как разницу между минимальными и максимальными значениями ЧОТ [F₀max-F₀min].

¹ Работа выполнена при финансовой поддержке Российского Научного Фонда (проект № 22-45-02007).

В кросс-культурном исследовании приняли участие 30 детей в возрасте 8-12 лет: 12 детей, для которых русский язык является родным (родились и живут в Санкт-Петербурге, Россия); 18 детей — носителей тамильского языка (родились и живут в Веллуру, Индия). В перцептивном эксперименте в качестве экспертов выступали взрослые с профессиональным опытом в области речевых исследований, говорящие на русском и тамильском языках (по 10 взрослых). По спонтанной речи русскоязычных детей русские эксперты лучше определили состояние радости (60 % правильных ответов) и нейтральное состояние (84 %) по сравнению с состояниями печали (44 %) и гнева (25 %). Индийские эксперты хуже, чем российские эксперты, классифицировали состояния радости (39 %), печали (36 %) и нейтральное состояние (46 %) в речи русских детей, лучше - состояние гнева (46 %). Согласие между русскими и индийскими выявлено при распознавании состояния печали (полнота — 0,44 и 0,36). Русские эксперты определили эмоциональное состояние русских детей лучше, чем индийские ($p < 0,0001$ — тест Манна-Уитни) — особенно для нейтрального состояния ($p < 0,0001$) и состояния радости ($p < 0,01$), но не для печали и гнева.

Индийские эксперты определили нейтральное состояние (86 % правильных ответов), состояния печали (86 %) и гнева (81 %) в речи детей, говорящих на тамильском языке, лучше, чем состояние радости (80 %). Русские эксперты определили нейтральное состояние (72 %) и состояние печали (58 %) в речи детей, говорящих на тамильском языке, лучше, чем состояния радости (42 %) и гнева (42 %). Средняя полнота (UAR) определения эмоционального состояния по спонтанной речи российских детей для русских экспертов составила 0,53; для индийских экспертов — 0,42; UAR по спонтанной тамильской речи для русских экспертов — 0,54, для индийских экспертов — 0,83. Выявлено согласие между индийскими и русскими экспертами при определении нейтрального состояния по речи тамильских детей (полнота — 0,72 и 0,86). Индийские эксперты лучше, чем русские, распознали все эмоциональные состояния тамильских детей.

Таким образом, показано, что носители языка лучше распознают эмоции детей, говорящих на одном с ними языке, по характеристикам их речи: русские эксперты — русских детей, индийские эксперты — индийских детей. В тоже время индийские эксперты правильно определили эмоциональное состояние русских детей, русские эксперты индийских детей, но с меньшей вероятностью, чем детей, принадлежащих с ними к одной языковой среде. Межкультурные исследования показывают, что, хотя проявление базовых эмоций универсально [3], распознавание более точное, когда говорящие и воспринимающие (слушающие) принадлежат к одной культурной среде [4]. Насколько известно, кросс-культурные

данные о распознавании базовых эмоций русских и индийских детей (носителей тамильского языка) по речи носителями русского и тамильского языков отсутствуют.

Установлена связь между эмоциональным состоянием детей и акустическими характеристиками эмоциональной речи, правильно отнесенной русскими и индийскими экспертами к соответствующему эмоциональному состоянию (регрессионный анализ). Эксперты опираются на значения ЧОТ $\max F(1,35) = 4,264$ $p < 0,05$ ($R^2 = 0,109$ $\beta = -0,33$); и диапазон ЧОТ $[F0 \max - F0 \min]$ — $F(1,35) = 4,212$ $p < 0,05$ ($R^2 = 0,107$ $\beta = -0,33$).

Определены акустические характеристики речи, характеризующие эмоциональные состояния: Радость — интенсивность слов выше по сравнению с состоянием грусти и меньше по сравнению с гневом; минимальная длительность слов по сравнению со словами, отражающими грусть и гнев; значения ЧОТ слов выше, чем в нейтральном состоянии. Печаль — максимальная длительность слов, высокие значения ЧОТ (по сравнению с нейтральным состоянием), интенсивность речевых сигналов ниже, чем для состояний радости и гнева. Гнев - максимальные значения интенсивности речевых образцов, длительность фраз меньше, чем в состоянии радости.

Проведено исследование по распознаванию эмоционального состояния типично развивающихся детей (ТР) и детей с атипичным развитием (синдромом Дауна — СД и расстройствами аутистического спектра — РАС). Аудиторами явились 100 взрослых: студенты-педиатры — 50 человек, студенты-биологи — 50 человек. Перед аудиторами стояла задача определить эмоциональное состояние детей как «комфорт — нейтральное — дискомфорт» при прослушивании тестового материала. Результаты показали, что в анализируемых группах детей (ТР, РАС, СД) — эмоциональное состояние хуже определяют по речи ТР детей (комфорт — 58 % правильных ответов, дискомфорт — 56 %). Аудиторы лучше распознают состояние дискомфорта у детей с РАС и СД (78 % ответов), чем комфортное состояние (67 и 70 % — соответственно для детей с РАС и СД). Нейтральное состояние лучше распознают по речи детей с РАС (67 %), чем по речи детей с СД (52 %) и ТР детей (54 %).

Результаты автоматического распознавания по трем состояниям «комфорт — нейтральное — дискомфорт» на основе автоматически извлекаемых акустических признаков из набора eGeMAPSv01b показали, что средняя точность распознавания эмоциональных состояний для детей всех групп: набор GeMAPSv01b accuracy (точность) — 0,694; набор eGeMAPSv01b accuracy — 0,686. Средняя точность автоматического распознаваний эмоциональных состояний для детей с РАС: набор GeMAPSv01b accuracy — 0,687; набор eGeMAPSv01b accuracy — 0,671; для детей с СД: набор GeMAPSv01b accuracy — 0,725; набор

eGeMAPSv01b accuracy — 0,717; для детей ТР: набор GeMAPSv01b accuracy — 0,641; набор eGeMAPSv01b accuracy — 0,631. В целом эмоциональное состояние детей с РАС и СД распознается с большей точностью и полнотой, по сравнению с эмоциональным состоянием ТР детей. Сравнение распознавания эмоционального состояния детей человеком и машиной показало сходную зависимость в отношении распознавания комфортного состояния: у ТР детей определяется хуже, чем у детей с РАС и СД. В отношении состояния дискомфорта — в перцептивном эксперименте лучшее распознавание было по речи детей с РАС и СД, в автоматическом — по речи ТР детей.

Проведен анализ влияния организации тестовых последовательностей и индивидуальных особенностей аудиторов на распознавание эмоционального состояния детей. Показано, что определение эмоционального состояния ТР детей связано с заданием для аудиторов $F(5,37) = 28,773$ $p < 0,0001$ ($R^2 = 0,795$ $\beta = -0,573$) (чем больше заданий должны выполнить аудиторы, тем хуже распознавание) и количеством аудиторов, прослушивающих тесты ($\beta = -0,775$). На распознавание эмоционального состояния детей с РАС оказывает влияние $F(7,25) = 881,38$ $p < 0,001$ ($R^2 = 0,996$) количество речевых сигналов в тесте ($\beta = 0,295$), речевой материал ($\beta = -0,803$) (по вокализациям распознавание состояния осуществляется лучше, чем по словам и фразам), опыт аудитора ($\beta = 3,914$) — множественный регрессионный анализ. Факторы, влияющие на распознавание эмоционального состояния по речи детей с СД, не выявлены.

Полученные данные об отражении эмоциональных состояний в характеристиках голоса и речи детей в совокупности с баллами за выполненные ими задания по методике CEDM могут быть использованы для оценки сформированности эмоциональной сферы, особенно детей с атипичным развитием, что важно при выборе стратегий для обучения детей, при подборе персонала, работающего с детьми.

Список использованной литературы

1. *Lyakso E., Frolova O., Kleshnev E., Ruban N., Mekala M., Arulalan K. V.* Approbation of the Child's Emotional Development Method (CEDM) // In Companion Publication of the 2022 International Conference on Multimodal Interaction (ICMI '22 Companion). — New York, NY, USA, 2022. — Pp. 201–210.
2. *Lyakso E., Frolova O., Kaliyev A., Gorodnyi V., Grigorev A., Matveev Yu.* AD-Child.Ru: Speech Corpus for Russian Children with Atypical Development // LNAI 11658. — 2019. — Pp. 299–308.
3. *Ekman P.* Basic emotions. Handbook of cognition and emotion / Dalglish T., Power M. J. (eds). — New Jersey: John Wiley & Sons, Ltd, Hoboken, 1999.
4. *Laukka P., Elflein H. A.* Cross-cultural emotion recognition and in-group advantage in vocal expression: A meta-analysis // Emotion Review. — 2020. — Vol. 13, is. 1. — Pp. 3–11.

СОВРЕМЕННЫЕ ПРОБЛЕМЫ АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ РЕЧИ: ВЫЯВЛЕНИЕ И АНАЛИЗ ЭКСТРАЛИНГВИСТИЧЕСКИХ ВОКАЛИЗАЦИЙ В СПОНТАННОЙ РАЗГОВОРНОЙ РЕЧИ¹

Анастасия Андреевна Поволоцкая

Алексей Анатольевич Карпов

(ФГБУН «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН))

Введение. Спонтанная разговорная речь звучит везде: в телефонных разговорах, на улице, в аудиториях, видеоконференциях, среда звучания спонтанной речи везде разная, в результате чего, на речевой сигнал накладываются шумы окружения (аддитивный шум, реверберация), также, на качество речевого сигнала влияет звукозаписывающее оборудование. Не стоит исключать присутствие в спонтанной разговорной речи хезитационных явлений, вокализаций и ошибок, которые реализует говорящий, при формулировании высказывания. Подобные явления воспринимаются как: «лингвистический мусор» или артефакты, однако, выявление и анализ подобных вокализаций может помочь при проектировании систем для распознавания как физического, так и психоэмоционального состояния говорящего. На данный момент спонтанная разговорная речь представляет сложность для систем автоматического распознавания речи (далее АРР), связано это с вариативностью и неоднородностью в речевых событиях.

Обобщенный метод автоматического распознавания речи (см. рис. 1) состоит из следующих этапов, перечисленных ниже.

1. В первую очередь производится предобработка речевого сигнала: шумоподавление и использование алгоритма для поиска голосовой активности в речевом сигнале [1, 5, 11].
2. В дальнейшем осуществляется нарезка сигнала перекрывающимися сегментами (ширина окна = 20-30мс и шаг 10-15мс) и выделение низкоуровневых дескрипторов (либо выбирается определенный перечень из нескольких, либо предпочтение отдается наборам, таким, как ComParE или GeMAPS и др.) [3].
3. Акустическая модель на основе выделенных признаков выделяет фонемы/графемы в сигнале [12].

¹ Данное исследование выполняется в рамках Ведущей научной школы РФ (грант № НШ-17.2022.1.6).

4. Языковая модель осуществляет постобработку извлеченных фонем/графем из сигнала (мешки слов, n-граммы и др.) [12].



Рис. 1. Упрощенный базовый метод обработки речевого сигнала для последующего анализа экстралингвистических вокализаций

Количество исследований по распознаванию разговорной спонтанной речи ограничено в первую очередь небольшим количеством открытых наборов с данным типом речи, а, во-вторых, как правило при делении записей на файлы определенной длины, нарезка речевых сигналов осуществляется именно на моментах с ошибками и вокализациями, реализуемыми диктором.

Главную сложность в выявлении и анализе экстралингвистических вокализаций представляет нахождение тех экстралингвистических вокализаций (далее ЭВ), которые могут оказывать влияние на лингвистическое содержание высказывания, поскольку не все ЭВ характеризуют психоэмоциональное и физическое состояние говорящего. В качестве первостепенного этапа в вопросе решения задачи выявления ЭВ в разговорной звучащей речи можно поставить цель - выделение системой APR всех реализаций ЭВ.

Изучение ЭВ на данный момент осуществляется путем классификации отдельно записанных звуковых файлов с помощью различных алгоритмов машинного обучения и выделенных акустических признаков (MFCC, F0, HNR, LPCC и др.) [13]. Возможная причина недостатка должного внимания к изучению ЭВ, их выявлению и анализу с использованием технологий APR связано с отсутствием как определенного понимания зачем подобные речевые явления фиксировать, так и систематизации ЭВ.

Систематизация ЭВ. В лингвистике нет четкого определения, которое обобщало бы такие невербальные речевые явления, как плач, смех, вздохи/выдохи, кашель, щелканье и цыканье языком. Основываясь на проанализированной литературе [2, 12, 17, 18], выдвинуто предположение, что подобные явления целесообразно обозначить общим термином «экстралингвистические вокализации», поскольку они не являются вербальной составляющей речи, но присутствуя в ней, дополняют ее и несут дополнительный смысл.

Создание универсальной систематизации ЭВ невозможно, поскольку языки различаются по своей фонетической и лексической структурам. Так, например, щелчки, будучи не интегрированными в морфологическую и синтаксическую структуры языка, являются единицей для передачи эмоций (возмущения или раздражения). Однако, в таких языках, как иврит, аварский, сомалийский и некоторые другие, щелчки являются лексическими единицами [8, 14].

1. Врожденными ЭВ являются звуки, которые ребенок «осваивает самостоятельно» в результате развития функций, формирующиеся на этапах роста плода и первых дней жизни, к ним относятся следующие категории звуков:

1.1 Вегетативные звуки (физиологические) — дыхание, отрыжка, чихание, кашель, икота, зевота и др. вегетативные звуки не «изучаются», не «заучиваются». Однако существуют вегетативные звуки, которые требуют определенного уровня «научения», например, чмоканье губами или сморкание носом. Некоторые вегетативные звуки могут быть использованы намеренно, например, прочищение горла («кхм-кхм») для обозначения; «я уже здесь» или сигнал собеседнику, что о чем-то нужно промолчать. Таким образом, преднамеренное использование вегетативных звуков в речи требует определенного прагматического знания и контроля над речевым аппаратом.

1.2 Рефлекторные звуки — плач, смех, звуки выражения беспокойства и т.п. Подобные звуки сигнализируют об эмоциональном состоянии ребенка на ранних стадиях его развития, имеют социальную функцию. Позднее, по мере взросления, ребенок учится контролировать свои эмоции, например, сдерживать плач/смех и данные вокализации переходят в раздел индивидуальных особенностей речепорождения, поскольку по тем или иным реакциям можно будет составить психоэмоциональный и речевой портрет говорящего.

1.3 Протофоны (речеподобные звуки или свободные вокализации) — квазигласные и полные гласные, визг и рычание, шепот и крик, ингрессивные и эгрессивные вокализации. Воспроизводятся эндогенно при отсутствии внешних провоцирующих факторов. Не несут социальной функции, сигнализируют о психофизиологическом состоянии довербального ребенка. Рассматриваются, как этап прелингвистического развития и предпосылкой формирования речи [2, 4]. Позднее, с освоением речи данное явление реформируется в речь.

2. Приобретенные ЭВ — вокализации, которым ребенок научается в процессе социально-коммуникативного развития и реализует в своем речевом поведении. В данную группу вошли следующие вокализации.

2.1 Аффективы (междометия): «Ба!», «Ой!», «Ого!», «О-о-о!» и др., отражающие безадресность сообщения. Среди междометий существуют

различные степени лексикализации: «*Бляха-муха!*» или «*Черт!*» — это явно, вербальные вокализации, в то время как «*ух*» или «*дзинь*» кажутся менее конвенционализированными. Некоторые междометия являются аффективными словами с неграмматической фонологией, такими как «*тссс*» или «*шшиш*» (без гласных) и «*ц-ц-ц-ц*» (щелчки) [17, 18].

2.2 Эксплетивные единицы: употребление слов-паразитов: «*типо*», «*которое*», «*ну*», «*что это*» и др. также представляющие собой десемантизированные элементы.

2.3 Звуки нефонемного характера, частично или полностью «десемантизированные» элементы: «*э-э*», «*м-м*», и др., которые возможно расценивать, как проявления пауз-хезитаций; вопросно-ответные лексеммы такие, как: «*ага*», «*угу*», «*хм*», «*ммм?*» и др.

2.4 Артикулируемые и вокализованные явления: цыканье при выражении недовольства или раздражения, скрипения зубами, чмоканье губами (имитация поцелуя). Такие рефлекторные явления, как смех, плач, крик становятся подвластны человеку с точки зрения осознанного контроля, и личность использует их в рамках повседневной коммуникации с целью передачи прагматических сигналов [6, 17].

3. Группу индивидуальных особенностей речепорождения представляют следующие вокализации.

3.1 Частотное употребление отдельных вокализаций, которые составляют речевой портрет говорящего и отличают речь отдельного индивида.

3.2 Заикание, проявляющееся в пролонгации либо частом повторении отдельных звуков.

3.3 Тикоподобные вокальные и артикулируемые особенности: частотное причмокивание, покашливание, щелканье языком, и другое.

3.4 Особенности порождения речи при таких заболеваниях как, черепно-мозговые травмы, перенесенные психоэмоциональные потрясения и заболевания.

Группы заикания и тикоподобных особенностей составляют вокализации, которые носят врожденный (например, наследственные заболевания) либо приобретенный характер (психологическая травма, перенесенное заболевание).

На основе анализа литературы предлагается оригинальная систематизация ЭВ для русского языка, представленная на рисунке 2.

Метод обработки речевого сигнала для выявления ЭВ. В результате выделения перечня вокализаций, являющихся предметом поиска при АРР и анализа литературы по соответствующей тематике [3, 7, 9, 10, 15, 19], мы предлагаем следующий метод обработки речевого сигнала, который позволит в дальнейшем осуществлять анализ экстралингвистических вокализаций (см. рис. 2).

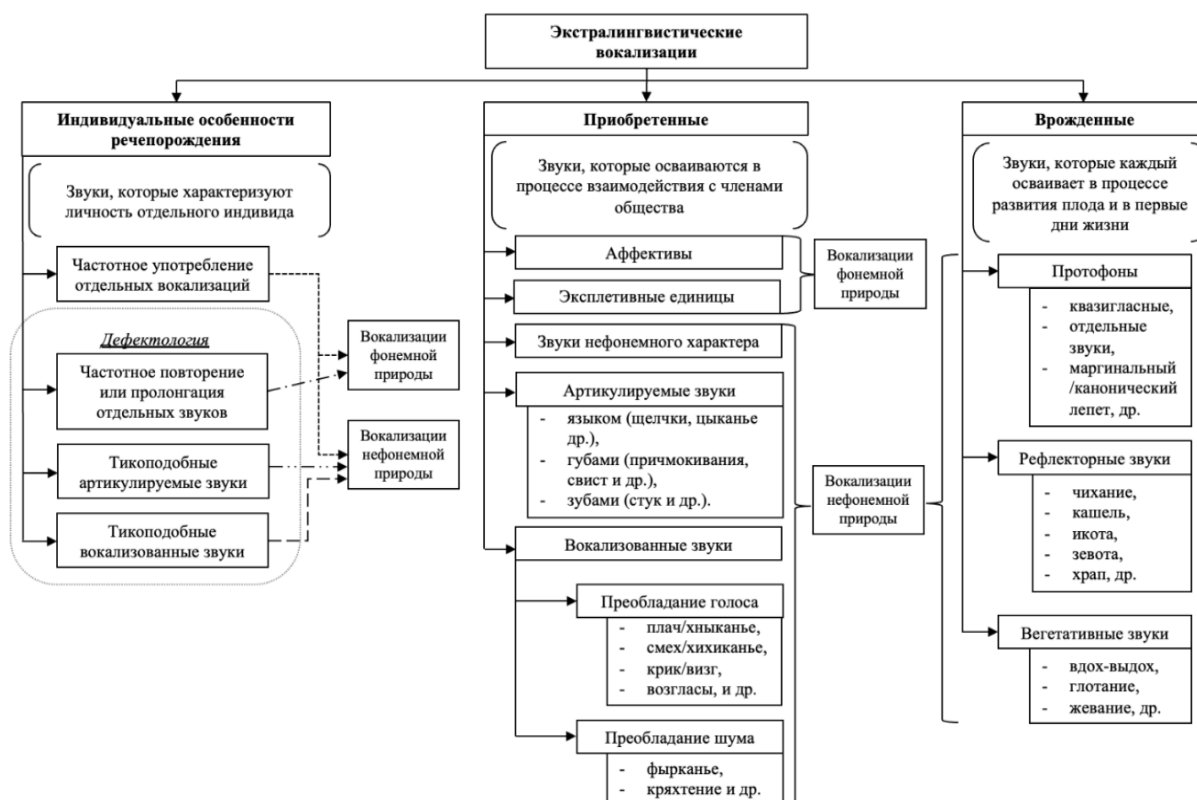


Рис. 2. Систематизация экстралингвистических вокализаций в русском языке

Поскольку спонтанная разговорная речь может звучать в любых условиях с разной степенью зашумленности и различным количеством говорящих, в первую очередь необходимо провести нормализацию сигнала и шумоподавление. В блоке шумоподавления, возможно, является целесообразным реализовать алгоритм CASA (Computational Auditory Scene Analysis, компьютерный анализ аудиальных сцен) на основе кохлеограммы или использовать искусственную нейронную сеть на основе конкатенации высокоуровневых признаков кохлеограммы и мел-спектрограммы.

В качестве входных данных для акустической модели предлагается использовать несколько видов изображений: спектрограммы, мел-спектрограммы и кохлеограммы сигнала. Кохлеограмма в отличие от спектрального представления сигнала, полученного методом скользящего окна, более точно отражает динамику сигнала во временной области, но более размыто в частотной, что позволит отслеживать изменения сигнала во времени.

Для выявления ЭВ извлечению подлежат такие низкоуровневые дескрипторы, как HNR, ZCR, MFCC (40 первых значений гребенки фильтров, дельта и дельта-дельта признаки), F0, и ряд других, которые будут подаваться на вход предобученной модели нейронной сети на материале, содержащем аудиофайлы реализаций ЭВ. Возможно, стоит произвести

аугментацию данных и наложить некоторые виды ЭВ на речь, например, смех, чтобы получилась реализация речевого смеха.

Предлагается использовать две различные языковые модели для вокализаций фонемной и нефонемной природы. Для вокализаций фонемной природы стоит использовать языковую модель на основе фонем, в случае с вокализациями нефонемной природы, стоит либо разработать символьную систему знаков, либо маркировать реализации подобных явлений обычными подписями, например, «кашель», «чихание», «икота» и др.

В дальнейшем осуществляется запись отдельных файлов для орфографической транскрипции текста и ЭВ вокализаций. В последствии возможна конкатенация разметок.

Заключение. В результате проведенного анализа актуальной литературы по теме выявления и анализа невербальных вокализаций была составлена оригинальная систематизация ЭВ и предложен метод обработки сигнала спонтанной разговорной речи.

На основе количественного и качественного анализа ЭВ, порождаемых диктором, можно сделать соответствующие выводы не только о его психоэмоциональном и физическом состоянии, но и также использовать методику выявления и анализа ЭВ при идентификации диктора, и распознавании звуков-команд для людей с моторно-речевыми нарушениями.

Список использованной литературы

1. Белов С. В., Катунин С. С. Гибридная методика шумоподавления в речевом сигнале для системы видеоконференций // Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная техника и информатика. — 2023, № 1. — С. 36–42. <https://doi.org/10.24143/2072-9502-2023-1-36-42>.
2. Леонтьев А. Н. Психологические основы дошкольной игры // Избранные психологические труды. — М.: Наука, 1983. — С. 303–323.
3. Обухов Д. С. Разработка современной системы распознавания русскоязычной телефонной речи // УБС. — 2021, вып. 89. — С. 106–122. <https://doi.org/10.25728/ubs.2021.89.4>.
4. Савельева Н. А. Особенности довербальной вокализации у плодов и младенцев // Специальное образование. — 2022, № 2 (66). — С. 246–259.
5. Тепляков А. Б., Спицын В. Г. Алгоритм обнаружения речевой активности в акустическом сигнале с применением свёрточных нейронных сетей // Молодежь и современные информационные технологии: сборник трудов XVII Международной научно-практической конференции студентов, аспирантов и молодых учёных, 17–20 февраля 2020 г., г. Томск. — Томск: Изд-во ТПУ, 2020. — С. 134–135.
6. Buder E. H. An acoustic phonetic catalog of prespeech vocalizations from a developmental perspective // Comprehensive perspective on child speech development and disorders: Pathways from linguistic theory to clinical practice. — Hauppauge, NY: NOVA, 2013.

7. *Dumpala S. H., Alluri K. N. R. K. R.* An Algorithm for Detection of Breath Sounds in Spontaneous Speech with Application to Speaker Recognition // Karpov A., Potapova R., Mporas I. (eds) *Speech and Computer. SPECOM 2017. Lecture Notes in Computer Science*. — 2017, vol. 10458. Springer, Cham. https://doi.org/10.1007/978-3-319-66429-3_9.
8. *Gil D.* Para-Linguistic Usages of Clicks // *The World Atlas of Language Structures Online*. 2013. [Электронный ресурс]. <https://doi.org/10.5281/zenodo.7385533>.
9. *Hsu J.-H., Su M.-H., Wu C.-H., Chen Y.-H.* Speech Emotion Recognition Considering Nonverbal Vocalization in Affective Conversations // *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. — 2021, vol. 29. — Pp. 1675–1686. <https://doi.org/10.1109/TASLP.2021.3076364>.
10. *Huang K.-Y., Wu C.-H., Hong Q.-B., Su M.-H., Chen Y.-H.* Speech Emotion Recognition Using Deep Neural Network Considering Verbal and Nonverbal Speech Sounds // *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. — Brighton, UK, 2019. — Pp. 5866–5870. <https://doi.org/10.1109/ICASSP.2019.8682283>.
11. *Ivry A., Berdugo B., Cohen I.* Voice activity detection for transient noisy environment based on diffusion nets // *IEEE Journal of Selected Topics in Signal Processing*. — 2019, vol. 13(2). — Pp. 254–264. <https://doi.org/10.1109/JSTSP.2019.2909472>.
12. *Kadali D. B., Kumar M. V.* Studies on Paralinguistic Speech Sounds // *IEEE 17th India Council International Conference (INDICON)*. 2020. <https://doi.org/10.1109/indicon49873.2020.9342586>
13. *Kumar Y., Mahajan M.* Machine learning based speech emotions recognition system // *International Journal of Scientific & Technology Research*. — 2019, vol. 8 (7). — Pp. 722–729.
14. *Ogden R.* Clicks and percussives in English conversation // *Journal of the International Phonetic Association*. 43. Pp. 299–320. <https://doi.org/10.1017/S0025100313000224>.
15. *Salamin H., Polychroniou A., Vinciarelli A.* Automatic detection of laughter and fillers in spontaneous mobile phone conversations // *IEEE SMC 2013: IEEE International Conference on Systems, Man, and Cybernetics, Manchester, UK, 13–16 Oct 2013*. — Pp. 4282–4287. <https://doi.org/10.1109/SMC.2013.730>.
16. *Serrurier A., Neuschaefer-Rube C., Röhrig R.* Past and Trends in Cough Sound Acquisition, Automatic Detection and Automatic Classification: A Comparative Review. — 2022, vol. 22. — P. 2896. <https://doi.org/10.3390/s22082896>.
17. *Trouvain J.* Laughing, breathing, clicking the prosody of nonverbal vocalizations // *Proc. Speech Prosody*. — 2014. — Pp. 598–602. <https://doi.org/10.21437/SpeechProsody.2014-108>.
18. *Trouvain J., Truong K. P.* Comparing non-verbal vocalizations in conversational speech corpora // *Proceedings of the 4th International Workshop on Corpora for Research on Emotion Sentiment and Social Signals*. — 2012. — Pp. 36–39.
19. *Zhu G., Yan Y., Caceres J.-P., Duan Z.* Transcription free filler word detection with Neural semi-CRFs // *Electrical Engineering and Systems Science, Audio and Speech Processing*. 2023.

О НЕЙРОННОЙ ОСНОВЕ ШКАЛ ОЦЕНКИ КОРОТКИХ ИНТЕРВАЛОВ ВРЕМЕНИ

*Дарья Никитична Подвигина
(Санкт-Петербургский государственный университет,
Институт физиологии им. И. П. Павлова РАН)*

Восприятие окружающей действительности неотъемлемо включает в себя оценку временных параметров сигналов и событий, поэтому вопрос о природе субъективного отражения времени занимает важное место в ряду центральных проблем целого ряда научных дисциплин — от философии до нейробиологии. По мнению исследователей, наибольший интерес для изучения базовых нейрофизиологических механизмов отражения времени представляет исходный уровень формирования временных впечатлений, которое осуществляется в условиях восприятия очень коротких временных интервалов — длительностью до 2–3 с [15; 5; 7]. Рассматривая именно этот временной диапазон, можно говорить о «непосредственном восприятии времени» [4; 16]. Кроме того, множество аспектов поведения, в том числе распознавание устной речи, требуют от человека способности к адекватной оценке длительностей событий, протекающих именно в диапазоне от десятков миллисекунд до нескольких секунд.

В то же время, нейронные механизмы процесса восприятия временных интервалов, в особенности, коротких длительностей, изучены далеко не полно. Исследователи сходятся во мнении, что оценка интервалов различной длительности осуществляется в результате работы сложной системы отражения времени, основывающейся на импульсной активности нейронов различной структура мозга. Какие же именно структуры мозга вовлечены в реализацию функции восприятия времени? И каковы принципы функционирования системы восприятия времени?

В исследованиях, использующих различные методы нейровизуализации, в первую очередь, функциональную магнитно-резонансную томографию (фМРТ), приводится множество данных об участии тех или иных регионов мозга в процессе оценки времени. Обзор [11] суммирует эти данные, рассматривая три гипотетических системы восприятия времени: (1) Общий механизм оценки времени. Данные ряда психофизических и клинических исследований предполагают, что можно выделить какую-то одну ключевую структуру, такую как, например, мозжечок или базальные ганглии, которая задействована в широком диапазоне функций, связанных с оценкой времени. Однако фМРТ-исследования показали, что скорее такой общий механизм восприятия времени является распределенным и связан с активацией большой сети, включающей дополнительную моторную зону, теменную и префронтальную кору, а также базальные ганглии и мозжечок. (2) Механизм частично распределенной

функции. Эта модель предполагает, что временная оценка осуществляется благодаря совместному функционированию нескольких областей мозга. К ним относятся регионы, которые неизменно участвуют во временной обработке вне зависимости от контекста - основная сеть оценки времени, включающая дополнительную моторную область и базальные ганглии, — и структуры, которые активируются в зависимости от контекста. (3) Распределенный механизм оценки времени. Данные методов моделирования и микроэлектродной регистрации активности отдельных нейронов предполагают, что оценка времени является неотъемлемым свойством нейронных сетей мозга и, следовательно, нет специальной структуры или нейронной сети для восприятия времени.

К последней группе нейронных механизмов оценки времени относится предложенная Buonanno и Merzenich [9; 12] модель динамической сети, состоящей из популяции взаимосвязанных возбуждающих и тормозных нейронов. Модель основывается на зависимых от времени свойствах нейронов, таких как попарное синаптическое облегчение и медленный тормозный постсинаптический потенциал. Компьютерная проверка данной модели показала, что она способна выполнять задание по распознаванию длительностей, а экспериментальные данные подтвердили наличие в разных областях мозга нейронов, селективно реагирующих на длительности и интервалы из диапазона от десятков до сотен миллисекунд [8; 12]. В коре млекопитающих были зарегистрированы слуховые нейроны, чувствительные к звукам разной длительности и интервалам разной продолжительности, а также в целом к временной структуре слуховых стимулов [18; 20]. Кроме того, избирательные к длительности нейроны были обнаружены и в зоне VI зрительной коры кошки [10].

Это значит, что нейронные сети способны самостоятельно обрабатывать временную информацию в диапазоне от десятков до сотен миллисекунд, без привлечения «внутренних часов» или других сложных моделей восприятия времени. Тогда согласованная деятельность нейронов, избирательных к разным длительностям стимулов, может служить основой для работы распределенной сети оценки времени. Помимо этого, ряд исследований показывает, что субъективная оценка временных интервалов и восприятие текущего времени может осуществляться за счет количественной оценки нейронными сетями периодов осцилляций, генерируемых нейронами различных структур мозга [17; 13]. Осцилляции различной частоты тогда могли бы составлять шкалы разной размерности для оценки временных интервалов и длительностей в широком диапазоне значений — от десятков до нескольких секунд.

Еще один возможный нейронный механизм, обеспечивающий оценку коротких временных интервалов, предложен в работе [19].

Авторы обнаружили в медиальной фронтальной коре обезьян нейроны, частота импульсации которых изменялась в зависимости от того, интервал какой длительности воспроизводили в поведенческом эксперименте обученные животные. То есть нейроны оказались способны масштабировать во времени свои ответы для лучшего соответствия моторному действию.

Таким образом, оценка временных интервалов в нервной системе происходит за счет распределенных в структурах мозга нейронов, обладающих различными частотно-временными характеристиками либо избирательно реагирующих на сигналы разной длительности. При оценке времени их согласованная работа формирует шкалу нервной активности, на которой отображается длительность оцениваемого временного отрезка. Существование в сенсорных системах шкал нервной активности, отображающих ту или иную мерность сигнала, обсуждалось еще в работах У. Таннера и Дж. Светса (цит. по [2]). Позднее исследовались характеристики шкал оценки сенсорными системами человека сигналов различных модальностей. В литературе рассматривается два вида шкал — непрерывные и дискретные. Непрерывные шкалы были получены, например, для оценки высоты, громкости и продолжительности речевого сигнала [6], дискретные — для оценки положения точки на линии [3; 1]. Данные психофизических исследований [14] свидетельствуют о том, что шкалы оценки коротких временных интервалов являются дискретными, содержат ограниченное число градаций и масштабируются в зависимости от диапазона оцениваемых длительностей, что согласуется с рассмотренными выше характеристиками нейронных сетей оценки времени.

Список использованной литературы

1. *Зысин С. Л.* Оценка положения точки на линии // Исследование принципов переработки информации в зрительной системе: сборник статей. — Л.: Наука, 1970. — С. 129–141.
2. *Леушина Л. И.* Зрительное пространственное восприятие. — Л.: Наука, 1978. — 176 с.
3. *Павловская М. Б.* О механизмах оценки размера зрительных изображений у человека: автореф. дис. ... канд. биол. наук: 03.03.01. — Л., 1977. — 18 с.
4. *Фресс П.* Восприятие и оценка времени // Экспериментальная психология: вып. 6 / под ред. П. Фресс, Ж. Пиаже. — М.: Прогресс, 1978. — С. 88–130.
5. *Холодная М. А.* Психология интеллекта. Парадоксы исследования. — 2-е изд., перераб. и доп. — СПб.: Питер, 2002. — 272 с.
6. *Чистович Л. А., Кожевников В. А.* Восприятие речи // Вопросы теории и методов исследования речевых сигналов: сборник. — Л.: АН СССР, 1969. — С. 4–149.
7. *Шиффман Х. Р.* Ощущение и восприятие: пер. с англ. — 5-е изд. — СПб.: Питер, 2003. — 928 с.

8. *Buonomano D. V., Hickmott P. W., Merzenich M. M.* Context-sensitive synaptic plasticity and temporal-to-spatial transformations in hippocampal slices // *Proceedings of the National Academy of Sciences of USA*. — 1997. — Vol. 94, no. 19. — Pp. 10403–10408.
9. *Buonomano D. V., Merzenich M. M.* Temporal information transformed into a spatial code by a neural network with realistic properties // *Science*. — 1995. — Vol. 267, no. 5200. — Pp. 1028–1030.
10. *Duysens J., Schaafsma S. J., Orban G. A.* Cortical off response tuning for stimulus duration // *Vision research*. — 1996. — Vol. 36, no. 20. — Pp. 3243–3251.
11. *Merchant H., Harrington D. L., Meck W. H.* Neural basis of the perception and estimation of time // *Annual review of neuroscience*. — 2013, vol. 36. — Pp. 313–336.
12. *Paton J. J., Buonomano D. V.* The neural basis of timing: distributed mechanisms for diverse functions // *Neuron*. — 2018. — Vol. 98, no. 4. — Pp. 687–705.
13. *Podvigin N. F., Bagaeva T. V., Boykova E. V., Zargarov A. Z., Podvigina D. N., Poeppel E.* Three bands of oscillatory activity in the lateral geniculate nucleus of the cat visual system // *Neuroscience Letters*. — 2004, vol. 361. — Pp. 83–85.
14. *Podvigina D. N., Lyakhovetskii V. A.* Characteristics of the Perception of Short Time Intervals // *Neuroscience and Behavioral Physiology*. — 2011. — Vol. 41, no. 9. — Pp. 936–941.
15. *Poeppel E.* Time perception // *Handbook of Sensory Physiology*. — 1978, vol. 8. — Pp. 714–729.
16. *Poeppel E.* A hierarchical model of temporal perception // *Trends of Cognitive Science*. — 1997. — Vol. 1, no. 2. — Pp. 56–61.
17. *Poeppel E.* The measurement of music and the cerebral clock: a new theory // *Leonardo*. — 1989. — Vol. 22, no. 1. — Pp. 83–89.
18. *Sadagopan S., Wang X.* Nonlinear spectrotemporal interactions underlying selectivity for complex sounds in auditory cortex // *Journal of Neuroscience*. — 2009. — Vol. 29, no. 36. — Pp. 11192–11202.
19. *Wang J. et al.* Flexible timing by temporal scaling of cortical responses // *Nature neuroscience*. — 2018, vol. 21. — Pp. 102–110.
20. *Zhou X. et al.* Successive-signal biasing for a learned sound sequence // *Proceedings of the National Academy of Sciences*. — 2010. — Vol. 107, no. 33. — Pp. 14839–14844.

АНАЛИЗ ЧЕРЕЗ СИНТЕЗ: УСТАНОВЛЕНИЕ АКУСТИЧЕСКИХ КЛЮЧЕЙ ИРОНИИ С ПОМОЩЬЮ МОДИФИКАЦИЙ ЗВУКОВОГО СИГНАЛА

*Павел Анатольевич Скрелин
Ульяна Евгеньевна Кочеткова
Евдокимова Вера Вячеславовна
Дарья Дионисовна Новосёлова*

(Санкт-Петербургский государственный университет)

В исследованиях, которые в разные годы выполнялись и продолжают выполняться на кафедре фонетики СПбГУ, важную роль играет процедура ресинтеза речевого сигнала как с изменением свойств и даже заменой сегментных единиц, так и с изменением его просодических характеристик, включая полную замену (пересадку) мелодического, динамического и темпорального контуров [1–3; 12]. Процедура ресинтеза позволяет точно определить, какие параметры и в каких пределах отвечают за реализацию функциональных свойств языковых единиц в потоке речи, а также за их эмоциональную и экспрессивную окраску.

Технологически суть ресинтеза заключается в пересадке определенных просодических параметров (например, мелодического контура или длительности ударного слога) от одного стимула на другой. Использование данного экспериментального подхода позволяет исследователям решить широкий ряд как чисто теоретических, так и прикладных задач: от изучения или получения определенной эмоциональной окраски и модальности в речи [1; 4; 5; 8; 10; 12–14; 17; 19] до оценки иностранного акцента [6; 7; 9; 15; 18].

В настоящем исследовании технология ресинтеза используется для изучения перцептивной значимости акустических средств при выражении иронии. Для установления акустических ключей иронии и их перцептивной значимости были проведены эксперименты, включающие различные наборы модификаций. При этом целью настоящей работы явилась проверка значимости акустических характеристик иронии, установленных в ходе исследования корпуса иронической русской речи [11]. Данный корпус включает в себя пары омонимичных высказываний, включенных в контексты, предполагающие прочтение с иронией и без иронии. Акустический и статистический анализ корпуса показал, что иронические стимулы отличаются от соответствующих им неиронических стимулов увеличением длительности и интенсивности ударного гласного, а также изменением мелодического контура [16].

Для проверки индивидуальной и комплексной значимости данных акустических характеристик при восприятии наличия или отсутствия иронии в речи были проведены эксперименты, включающие различные модификации звукового сигнала, в том числе пересадку мелодического контура. На первом этапе исследования были проведены аудиторские эксперименты, в которых исходные целевые фрагменты без какого-либо маркера наличия или отсутствия иронии были представлены аудиторам. Участники эксперимента должны были соотнести звучащий фрагмент с одним из контекстов (предполагающим иронию или нейтральное прочтение), предъявленном на экране. Эксперименты проводились на платформе SoSciSurvey. Для модификаций были отобраны такие пары стимулов, в которых и ироничный, и неироничный стимул, произнесенные одним и тем же диктором, были правильно опознаны не менее 85 % аудиторов.

Далее были проведены две серии модификаций. В первой серии исходными стимулами явились нейтральные высказывания, на которые накладывались характеристики соответствующих им ироничных стимулов, произнесенных тем же диктором. Во второй серии, наоборот, реципиентами становились ироничные стимулы, на которые пересаживались характеристики нейтральных стимулов. В обеих сериях экспериментов были проведены семь типов фонетических модификаций: длительности ударного гласного; интенсивности ударного гласного; мелодического контура целевого фрагмента; длительности ударного гласного с последующим изменением интенсивности; длительности ударного гласного с последующим изменением мелодического контура; интонационного контура, сопровождающееся модификацией интенсивности ударного гласного; длительности ударного гласного с последующей модификацией мелодического контура и изменением интенсивности ударного гласного. Таким образом, из исходных 18 омонимичных пар стимулов получилось 288 модифицированных фрагментов, которые были представлены на аудиторскую проверку в 6 перцептивных экспериментах. Участникам эксперимента предлагалось прослушать запись отрывка и определить, из какой фразы или диалога он был взят — из иронического или неиронического контекста. В каждом эксперименте аудиторам были предъявлено по 48 фрагментов. Результаты подводились после того, как эксперимент проходило не менее 20 респондентов [1].

Результаты пересадки акустических характеристик от ироничных фрагментов на ироничные и, наоборот, от неироничных на ироничные позволяют сделать вывод о том, что наиболее важным акустическим параметром, отвечающим за ироническую оценку, является мелодическое оформление (включая комбинации изменения интонационного контура

с другими акустическими характеристиками). Одновременное изменение длительности, интенсивности и интонационного контура оказывает наибольшее влияние на восприятие фразы как иронической или нейронической (см. рис. 1). В то же время наименее важными являются темпоральные и динамические характеристики; изменение этих параметров по отдельности или в совокупности друг с другом не оказывает значительного влияния на интерпретацию ироничных и нейроничных отрывков.

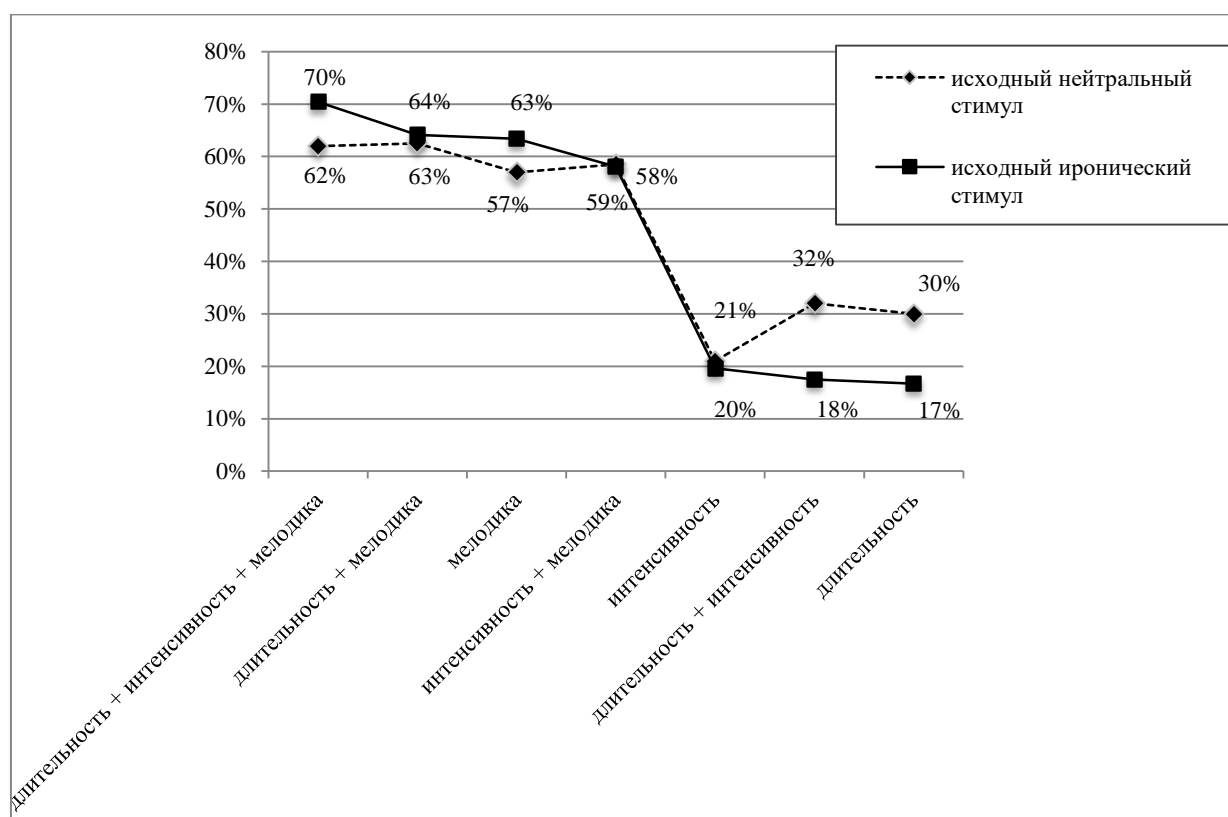


Рис. 1. Средний процент правильного распознавания стимулов как ироничных (на основе исходных нейтральных) и как нейтральных (на основе ироничных) в зависимости от типа модификаций и их комбинаций

Похожие результаты были получены и на материале других языков при пересадке характеристик иронических высказываний на нейронические, при этом на материале немецкого языка важным для восприятия иронии оказалось наличие дополнительной просодической выделенности, приводящей к изломанному характеру мелодического контура [4; 5], тогда как на материале французского языка [8] была доказана наибольшая эффективность одновременного изменения всех трех параметров: расширения диапазона частоты основного тона на один полутон, увеличения длительности стимула на 30% и изменения интонационного контура таким образом, чтобы он соответствовал конструкции Н+!Н* !Н%. Кроме того, было установлено, что во французском языке удлинение ударного слова и изменение интонационного контура в большей степени

отвечают за восприятие стимула как ироничного, чем увеличение диапазона частоты основного тона (подробный обзор зарубежных исследований представлен в работе [1]).

При этом можно заметить, что при пересадке характеристик иронических высказываний на нейронические, изменение сразу трех параметров приводит практически к такой же перцептивной оценке, что и изменение мелодики одновременно с длительностью. При обратной пересадке — от нейтральных высказываний на иронические — сочетание всех трех типов модификаций дает более высокий процент правильного опознание, тогда как перцептивная оценка изменений мелодического контура вместе с длительностью мало отличается от пересадки мелодического контура в отдельности.

Интересное явление наблюдалось при пересадке мелодического контура от нейтральных высказываний на иронические. Такие модификации приводили к изменению качества голоса, а также к эффекту назализации. Это связано с тем, что исходные иронические высказывания, в отличие от нейтральных, имели особую тембральную окраску: значения спектральной плотности в иронических высказываниях превышали аналогичные значения в нейтральных высказываниях. Поскольку на данный момент не существует методики модификации спектральной картины, которая позволила бы произвести идеальное удаление или пересадку данных акустических характеристик из одного звукового сигнала в другой, разработка подобных методов представляется перспективным и передовым направлением.

Список использованной литературы

1. *Новосёлова Д. Д.* Акустические характеристики иронии в русском языке. Выпускная квалификационная работа, уровень образования: магистратура, направление 45.04.02 «Лингвистика», Основная образовательная программа ВМ.5715. «Общая и прикладная фонетика», 2023.
2. *Морозова Н.А.* Сопоставительный анализ равнофункциональных интонационных контуров в разных языках. Выпускная квалификационная работа на степень бакалавра, 2017. <http://hdl.net/11701/7855>
3. *Сегаль Н. А.* Автоматическая стилизация мелодических контуров. Выпускная квалификационная работа, 1998.
4. *Burkhardt F., Weiss B.* Speech Synthesizing Simultaneous Emotion-Related States // *Proceedings of the International Conference on Speech and Computer (SPECOM 2018)*. — 2018, vol. 11096. — Pp. 76–85.
5. *Burkhardt F., Steinhilber A., Weiss B.* Ironic Speech — Evaluating Acoustic Correlates by Means of Speech Synthesis // *Proceedings of the 29th Conference on Electronic Speech Signal Processing (ESSV)*. — 2018. — Pp. 342–350.
6. *Felps D., Bortfeld H., Gutierrez-Osuna R.* Foreign accent conversion in computer assisted pronunciation training // *Speech Communication*. — 2009. — Vol. 51, no. 10. — Pp. 920–932.

7. *Felps D., Geng C., Gutierrez-Osuna R.* Foreign Accent Conversion through Concatenative Synthesis in the Articulatory Domain // *IEEE Transactions on Audio, Speech, and Language Processing*. — 2012. — Vol. 20, no. 8. — Pp. 2301–2312.
8. *González-Fuente S., Prieto P., Noveck I.* A fine-grained analysis of the acoustic cues involved in verbal irony recognition in French // *Proceedings of the 8th International Conference of Speech Prosody*. — 2016. — Pp. 902–906.
9. *Jügler J., Zimmerer F., Trouvain J., & Möbius B.* The Perceptual Effect of L1 Prosody Transplantation on L2 Speech: The Case of French Accented German // *Proceedings of the 17th Conference of the International Speech Communication Association (INTERSPEECH)*. — 2016. — Pp. 67–71.
10. *Kawahara H., Morise M.* Technical foundations of TANDEM-STRAIGHT, a speech analysis, modification and synthesis framework // *Sadhana*. — 2011, vol. 36. — Pp. 713–727.
11. *Kochetkova U., Skrelin P., Evdokimova V., Novoselova D.* The Speech Corpus for Studying Phonetic Properties of Irony // *Chernigovskaya T., Eismont P., Petrova T. (eds) Language, Music and Gesture: Informational Crossroads (LMGIC 2021)*, pp. 203–214. Springer, Singapore (2021). DOI: 10.1007/978-981-16-3742-1_16
12. *Kochetkova U., Skrelin P., Novoselova D., German, R.* Prosodic Features of Verbal Irony in Russian and French: Universal vs. Language-Specific. *Speech and Computer 24th International Conference, SPECOM 2022 Gurugram, India, November 14–16, 2022 Proceedings*, 13721, p. 358–371. Springer Verlag. DOI: 10.1007/978-3-031-20980-2_31
13. *Lei Y., Yang S., Wang X., Xie L.* *MsEmoTTS*. Multi-Scale Emotion Transfer, Prediction, and Control for Emotional Speech Synthesis // *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. — 2022, vol. 30. — Pp. 853–864.
14. *Moraes J.* The pitch accents in Brazilian Portuguese: analysis by synthesis // *Proceedings of the 4th International Conference of Speech Prosody*. — 2008. — Pp. 389–397.
15. *Rognoni L., Busà, M.* Testing the Effects of Segmental and Suprasegmental Phonetic Cues in Foreign Accent Rating: An Experiment Using Prosody Transplantation // *Proceedings of the International Symposium on the Acquisition of Second Language Speech (New Sounds 2013)*. — 2014, vol. 5. — Pp. 547–560.
16. *Skrelin P., Kochetkova U., Evdokimova V., Novoselova D.*: Can we detect irony in speech using phonetic characteristics only? — Looking for a methodology of analysis // *Karpov A., Potapova R. (eds.) SPECOM 2020. LNCS (LNAI)*, vol. 12335, pp. 544–553. Springer, Cham (2020). DOI: 10.1007/978-3-030-60276-5_52
17. *Tao J., Kang Y., Li A.* Prosody conversion from neutral speech to emotional speech // *IEEE Transactions on Audio, Speech, and Language Processing*. — 2006. — Vol. 14, no. 4. — Pp. 1145–1154.
18. *Vieru-Dimulescu B., Mareüil P.* The Contribution of prosody to the Perception of Foreign Accent // *Phonetica*. — 2006. — Vol. 63, no. 4. — Pp. 247–267.
19. *Wu X., Cao Y., Lu H., Liu S., Kang S., Wu Z., Liu X., Meng H.* Exemplar-Based Emotive Speech Synthesis // *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. — 2021, vol. 29. — Pp. 874–886.

АППРОКСИМАЦИЯ И ХЕДЖИРОВАНИЕ В ЯЗЫКЕ И РЕЧИ¹

Янань Сян

Наталья Викторовна Богданова-Бегларян
(Санкт-Петербургский государственный университет)

В лингвистике вопрос о соотношении *языка* и *речи* впервые был поставлен Ф. де Соссюром [12]. *Язык* — это символическая система, организованная фонологией как внешней средой, семантикой как внутренним значением, лексическим материалом и грамматическими правилами, а также правилами использования речи, разделяемыми членами общества в их речевой практике. *Речь*, в отличие от символической системы, представляет собой совокупность слов, которые люди фактически используют и производят во своем повседневном общении в качестве речевых актов.

Устная речь интересна для лингвистического исследования по многим причинам. В частности, в русском повседневном общении существует множество единиц, восходящих к обычным лексемам, как полнозначным, так и служебным, которые попадают под действие весьма активного процесса *прагматикализации* и в результате в ряде своих употреблений утрачивают (полностью или частично) лексическое и/или грамматическое значение и приобретают прагматическое, переходят в разряд коммуникативно-прагматических функциональных единиц [17; 16: 296]. Для них в коллоквиалистике используется термин *прагматический маркер* (ПМ) [1: 23].

Одним из типов ПМ являются *аппроксиматоры* (ПМА) — маркеры нечеткой, или приблизительной, номинации, показывающие неуверенность говорящего в том, о чем он говорит, или употребляющиеся, когда прямое название предмета, явления или положения дел является излишним, неуместным или невозможным [11: 632]. Типичными ПМА русской речи являются *типа, как бы, вроде, или там*. Все они достаточно частотны в устной коммуникации: так, маркер-аппроксиматор *как бы* имеет в частотном списке 60-ти базовых русских ПМ ранг 5 и IPM 900

¹ Исследование частично проведено при финансовой поддержке гранта Санкт-Петербургского государственного университета (проект № 94033528 «Моделирование коммуникативного поведения жителей российского мегаполиса в социально-речевом и прагматическом аспектах с привлечением методов искусственного интеллекта»).

(для 300 тыс. токенов в аннотированном подкорпусе ОРД¹), маркер *типа* — ранг 15 и ИРМ 297, маркер *вроде* — ранг 23 и ИРМ 77 [ПМ 2021: 54, 56]. Существенно менее частотным является ПМА *или там* (ранг 51 и ИРМ 3), но лишь потому, что его вычленение в ходе аннотирования корпусного материала сопряжено со значительными трудностями (см. об этом подробнее: [5; 10: 210; 13]).

Употребление ПМА — это частный случай *аппроксимации* в языке. В логике под аппроксимацией (от лат. *approximare* — ‘приближаться’) понимают «приблизительное выражение каких-либо величин через другие, более простые или более известные величины, в том или ином смысле близкие к исходным» [6: 48]. Аппроксимация, основанная на сравнении, является универсальной логической категорией (ср. упомянутые выше ПМА *типа, как бы, вроде*). В лингвистике не существует единства мнений относительно трактовки аппроксимации как языкового феномена, но обычно под *аппроксимацией* понимается именно приближительная номинация, т. е. неточное, приближительное именование различных объектов действительности: предметов, количеств, качеств, процессов [3: 100]. В частности, говорящий в устной коммуникации может испытывать трудности при поиске нужного слова или выражения, может вообще его не знать или может считать его по какой-либо причине нежелательным для употребления [10: 304].

Близким к *аппроксимации* является понятие *хеджирования* (от англ. *Hedge* — ‘уклонение от прямого ответа, страховка’) [18], т. е. использования слов (единиц) «осторожной» модальности — неопределенных местоимений или наречий, а также вводных слов, маркирующих неуверенность говорящего. Согласно теории речевых актов, стратегия хеджирования направлена на ослабление иллокутивной силы высказывания, так как хеджи придают высказыванию расплывчатость, частично снимают с говорящего ответственность за сказанное и смягчают категоричность его утверждений [15]. Можно сказать, что хеджи и маркеры-аппроксиматоры взаимно усиливают неуверенность говорящего, ср. обилие соответствующих единиц (выделены шрифтом) в небольшом контексте:

- 1) *А нет/ он **там** по-русски кусок и кусок по-английски/ я **такая типа**/ ну я **там** отрывисто **что-то** поняла/ что он **типа** едет брать интервью у Сталина/ **туда-сюда**/ и вот он хочет сва-*

¹ ОРД [9], или «Один речевой день» — это корпус русского языка повседневного общения, созданный в СПбГУ по методике 24-часовой записи (см. о нем, например: [2]) [<https://ord.spbu.ru>].

лить — вот посмотреть на этот голодомор/ он у него спрашивает/ как доехать или **что-то в этом духе** (УП)¹.

Интересно также отметить, что ПМА часто в русской речи выступают как полифункциональные единицы, т. е. реализуют, кроме аппроксимативной, еще функцию хезитации, навигатора по тексту, ксенопоказателя или ритмообразующую [10: 112, 206, 210, 396]. Реализовать эти функции часто помогают другие ПМ, образуя в речи разнообразные прагматические цепочки, ср.:

- 2) Ну я-я такая... я уже ищу — ну нету их вообще нигде/ я такая думаю/ типа/ ну/ я такая вылезая/ вижу эту тётку/ такая/ типа/ «Извините/ а где тут очки?» (УП);
- 3) Ну и здесь есть там всякие процедуры/ там не знаю/ бодания/ взаимного кусания/ да/ там не знаю/ какие-нибудь там тетерева или там козлы бодаются/ да/ там/ или ещё чего-нибудь (УП).

В приведенных примерах видно не только соседство в одном контексте ПМ разного типа, но и наличие хеджей (в контексте (3) подчеркнуты).

Близость аппроксимации и хеджирования как явлений устного дискурса поддерживается еще и наличием структурных вариантов базовых ПМ, в которых объединены ПМА и хеджи, ср.:

- 4) По аське / дословно не помню... **Что-то типа** / «Привет / как делишки?» (УП);
- 5) А... «честно говоря / я не столь тревожен / сколь правил тревогой» мне кажется / что он выражает не соотношение степени а... т-тревоги в указанной тревоге / а обозначает **что-то типа** / что я вообще не тревожился / а просто делал вид (УП);
- 6) [Слава] Макс. Максим. Надо слоган какой-то убедительный придумать. **Типа там** / Я дам вам... Что он им может дать? [Леша] Смотря чего у них нет [Слава] Ничего у них нет [Леша] Такой слоган / Я дам вам всё (УП).

В примерах (2), (4)–(6) такие структурные реализации (или цепочки разных ПМ — этот вопрос вряд ли имеет одно-единственное решение), как *такая типа*, *что-то типа*, *типа там*, выступают, с одной стороны, в роли аппроксиматора/хеджа, а с другой — в роли маркера-ксенопоказателя, который вводит в повествование *чужую речь* (в широком понимании данного термина) (в контекстах подчеркнута). Наличие подобных

¹ УП — это устный подкорпус Национального корпуса русского языка [8: <https://ruscorpora.ru>], основной источник материала для настоящего исследования.

маркеров показывает, что говорящий не уверен в точности передаваемых чужих слов (4), а иногда они вообще только предполагаются им (5)–(6).

На функцию аппроксиматора/хеджа может накладываться не только функция ксенопоказателя, но и другие, по преимуществу хезитационная, ср.:

- 7) *И приводит пример что-то типа/ он хочет открыть свой бизнес (УП);*
- 8) *И всякому/ конечно/ лингвисту-любителю очевидно/ что это город — ну/ что-то типа Львов/ то есть... Тут говорить не о чем/ совпадение полное/ с точностью до орфографической разницы в одной букве (УП);*
- 9) *Ну были/ конечно/ такие трудные институты/ это как бы серьёзные типа там/ я не знаю/ там где-то технические или там/ не знаю/ университет (УП);*
- 10) *Ну/ типа/ там/ «искривлённый»/ «учащённый»/ которые от возвратных глаголов/ да/ и частицы/ и вообще/ так сказать/ типа «заспанные глаза»/ «восторженный поклонник»/ «воздержанный образ жизни»/ которые вообще как бы не соответствуют в буквальном смысле никаким глаголам (УП).*

В контекстах (9)–(10) подчеркнуты и другие аппроксиматоры/хеджи, помимо выделенных, при этом все эти примеры — из речи весьма образованных носителей языка, в основном из научных лекций или докладов, что позволяет заключить, что употребление подобных единиц отнюдь не свидетельствует о низкой речевой компетенции говорящих. Скорее, аппроксимация и хеджирование — это универсальная черта устной коммуникации, свойственная к тому же общению на любом языке.

Есть уже и некоторые наблюдения над особенностями употребления аппроксиматоров/хеджей носителями языка с разным психотипом: так, оказалось, что и в диалоге, и в монологе экстраверты хеджируют в 2–4 раза реже, чем интроверты [4; 7; 14]. Это еще повышает значимость изучения подобных единиц в русской устной речи.

Новизна настоящего исследования заключается в комплексном анализе явлений аппроксимации и хеджирования, особенно маркеров-аппроксиматоров, в русской повседневной речи. В работе используются такие научные методы исследования, как метод сплошной выборки (для создания пользовательского подкорпуса материала), описательный (контекстный анализ), сопоставительный (данные словарей и устного корпуса) и количественный (простые количественные подсчеты).

Полученные результаты могут быть полезны как в коллоквиалистике (теоретический аспект), так и в практике преподавания русского

языка как иностранного и в практике перевода русских художественных текстов на другие языки, особенно на китайский (прикладные аспекты).

Список использованной литературы

1. *Богданова-Бегларян Н. В.* Предисловие редактора // Прагматические маркеры русской повседневной речи: словарь-монография / сост., отв. ред. и автор предисл. Н. В. Богданова-Бегларян. — СПб.: Нестор-История, 2021. — С. 5–52.
2. *Богданова-Бегларян Н. В., Шерстинова Т. Ю., Блинова О. В., Мартыненко Г. Я.* Корпус «Один речевой день» в исследованиях социолингвистической вариативности русской разговорной речи // Анализ разговорной русской речи (АР³-2017): Труды седьмого междисциплинарного семинара / науч. ред. Д. А. Кочаров, П. А. Скредин. — СПб.: Политехника-принт, 2017. — С. 14–20.
3. *Бузаров В. В., Лынова Э. Г.* Что такое аппроксимация в лингвистике? // Иностранные языки в школе. — 1991, № 2. — С. 100–102.
4. *Горбунова Д. А.* Прагматические маркеры русской устной речи: корреляция с психотипом говорящего. Дис. ... канд. филол. наук. — СПб., 2021. — 138 с. (машинопись).
5. *Зайдес К. Д.* Прагматический маркер *ИЛИ ТАМ*: свой среди чужих, чужой среди своих // Русская речь. — № 1, 2021. — С. 22–36.
6. *Кондаков Н. И.* Логический словарь-справочник. — М.: Наука, 1978. — 720 с.
7. *Костина Е. Д.* «Осторожные» интроверты и экстраверты: анализ способов хеджирования на материале монологов-рассказов об отдыхе // Компьютерная лингвистика и интеллектуальные технологии. По м-лам ежегодной международной конференции «Диалог» (2022). Мат-лы студенческой сессии / гл. ред. В. П. Селегей. — М.: РГГУ, 2022. — С. 1–5.
8. НКРЯ — Национальный корпус русского языка. <https://ruscorpora.ru/>
9. ОРД — Корпус русской повседневной речи «Один речевой день». <https://ord.spbu.ru/>
10. ПМ — Прагматические маркеры русской повседневной речи: словарь-монография / сост., ред. и автор предисл. Н. В. Богданова-Бегларян. — СПб.: Нестор-История, 2021. — 520 с.
11. *Подлеская В. И.* Нечеткая номинация в русской разговорной речи: опыт корпусного исследования // Компьютерная лингвистика и интеллектуальные технологии. По материалам ежегодной Международной конференции «Диалог» (2013). Вып. 12 (19). В двух томах. Том 1. Основная программа конференции / гл. ред. В.П. Селегей. — М.: РГГУ, 2013. — С. 631–643.
12. *Соссюр Ф. де.* Курс общей лингвистики / ред. Ш. Балли и А. Сеше; перев. с франц. А. М. Сухотина. — М.: Соцэкгиз, 1933. — 272 с.
13. *Сян Янань.* Прагматический маркер-аппроксиматор ИЛИ ТАМ: корпусное исследование // Тезисы 51-й Международной научной филологической конференции имени Людмилы Алексеевны Вербицкой. — СПб., Изд-во СПбГУ, 2023а. — С. 1015–1016.
14. *Сян Янань.* Насколько мы готовы отвечать за свои слова? (прагматические маркеры-аппроксиматоры в русском устном дискурсе) // Когнитивные исследования языка. Когнитивная лингвистика в контексте современной науки. — Челябинск, 2023б. — В печати.
15. *Fraser B.* Forthcoming. “A Brief History of Hedging” // *Vagueness in Language*, S. Schneider (ed.). — Bingley: Emerald Publishing, 2013. — Pp. 201–213.

16. *Graf E.* Interjektionen im Russischen als Interaktive Einheiten. — Frankfurt am Main, 2011. — 328 p.
17. *Günthner S., Mutz K.* Grammaticalization vs Pragmaticalization? The Development of Pragmatic Markers in German and Italian // W. Bisang, N. P. Himmelmann, B. Wiemer (eds). What Makes Grammaticalization? A Look from its Fringes and its Components. — Berlin: Language Arts & Disciplines, 2004. — Pp. 77–107.
18. *Lakoff G.* Hedges: A Study in Meaning Criteria and the Logic of Fuzzy Concepts // Journal of Philosophical Logic. — No 2. Dordrecht-Holland: D. Reidel Publishing Company, 1973. — Pp. 458–508.

АВТОМАТИЧЕСКАЯ КЛАСТЕРИЗАЦИЯ ТОНАЛЬНЫХ АКЦЕНТОВ В РУССКОМ ЯЗЫКЕ

Елизавета Владимировна Фаненштыль
(*Московский государственный университет им. М. В. Ломоносова,*
НИЦ «Курчатовский институт»)

Акцент — просодическое явление, при котором некоторый сегмент в синтагме получает просодическую выделенность, или проминентность [2]. Акцентирование может осуществляться посредством разных акустических параметров, основными являются частота основного тона, длительность и интенсивность. Как отмечает С. Оде, «<...> в русском, как и в голландском и английском языках, параметр высоты тона — самый сильный для передачи выделенности» [2: 3]. Таким образом, в русском языке мы чаще всего имеем дело с *тональным акцентом*.

Движение тона в акценте несет важную семантическую нагрузку — как писал С. В. Кодзасов, «наиболее важные функции интонации выполняются акцентами» [1: 382]. Например, оно передает незавершенность/завершенность фразы, или, иначе говоря, «ожидаемость продолжения» [2: 310]. Кроме того, оно кодирует иллокутивную цель высказывания: базовое представление о тонах состоит в том, что восходящие акценты используются в вопросах, а нисходящие — в утверждениях [там же]. В некоторых работах предлагается классификация акцентов в соответствии с функциями, например у Т. М. Николаевой встречаем «фразовое ударение, контрастное парадигматическое, контрастное синтагматическое, выделение темы, подчеркивание» [6: 139]. Часто акцентов в синтагме несколько, один из них принято считать главным (несущим/ядерным). Главный акцент обычно располагается в конце синтагмы и является самым интенсивным.

Ключевой задачей работы является определить основные типы несущего тонального акцента в современной русской речи и коммуникативные типы высказываний, с которыми они сопряжены. Большинство известных работ, ставящих целью выявить интонационные паттерны в русской речи, применяют интроспективный подход: тональные движения часто анализируются скорее на основе слухового впечатления, чем на основе акустических данных. Мы поставили перед собой цель как можно более автоматизировать процесс просодического исследования, поэтому для вывода искомым тональных типов мы применяем кластерный подход.

Кластеризация — алгоритм, который разбивает множество контуров на однородные непересекающиеся подмножества. Данный алгоритм

интересен тем, что позволяет полуавтоматически обработать большой объем данных и исключает возможную пристрастность к объекту изучения. В последнее время кластерный анализ в применении к фонетике и интонации набирает популярность: методы кластеризации применялись в системах автоматического распознавания по голосу [9], также кластерный анализ тональных контуров применялся для проверки правильности предполагаемых интонационных категорий [11], наконец, при помощи кластерного подхода предпринимались попытки выявить тональные паттерны — например, для немецких пограничных тонов [10].

Анализ проводился на базе корпуса устной неподготовленной речи «Рассказы и разговоры о грушах» [4]. Материалом исследования служили аудиозаписи 6-ти говорящих (3-х мужчин и 3-х женщин в возрасте от 20-ти до 30-ти лет без речевых расстройств), и их вокальные аннотации в формате .textgrid и .doc. На основе этой разметки после дополнительной проверки в программе PRAAT были сегментированы несущие акценты; для соблюдения единообразия было решено исследовать сегмент, состоящий из предупредительного и ударного слога, ассоциированного с акцентом. Всего было выделено 408 контуров.

Для осуществления алгоритма мы использовали программу «Contour Clustering», предложенную кельнским фонетистом К. Каландом [3]. Перед кластеризацией производился набор измерений ЧОТ на множестве тональных контуров и вычислялась матрица расстояний между векторами измерений. Для определения расстояния мы выбрали алгоритм динамической трансформации временной шкалы, так как он позволяет учесть временное несовпадение тональных изменений, являющихся похожими по форме, но реализованными на разных временных отрезках. Поскольку тональный диапазон у разных говорящих отличается, мы также выполнили стандартизацию ЧОТ: для каждого говорящего средняя ЧОТ вычиталась из каждого её измерения, разность делилась на стандартное отклонение ЧОТ. Кроме того, была проведена конвертация ЧОТ, измеренной в Гц, в полутона: полутоновая шкала является логарифмической и более точно соответствует восприятию высоты тона, чем шкала Герца [7].

В качестве алгоритма кластеризации мы использовали иерархический аггломеративный подход, так как эта разновидность кластерного анализа не требует априорного указания количества кластеров и поэтому является оптимальной для визуального анализа данных. В начале работы данного алгоритма все контуры являются отдельными кластерами, на последующих шагах более мелкие кластеры объединяются до тех пор, пока их число не станет равным 2. В качестве метода определения расстояния мы задали метод Уорда, который позволяет поддерживать наименьшую внутрикластерную дисперсию.

Вычисление оптимального количества кластеров выполнялось программе R, версия 4.0.2 [8] с помощью функции NbClust пакета NbClust [12]. Данная функция предоставляет возможность применить множество статистических тестов — в том числе распространенный «метод локтя». Бóльшая часть тестов показала, что наилучшее количество кластеров — 5. В дополнение к указанной функции также были применены тесты, предложенные в программе «Contour Clustering»: метод, основанный на принципе минимальной длины описания, и метод, вычисляющий внутригрупповую и межгрупповую дисперсии. Эти тесты также показали результат, равный 5.

Результаты кластеризации приведены на рисунке 1.

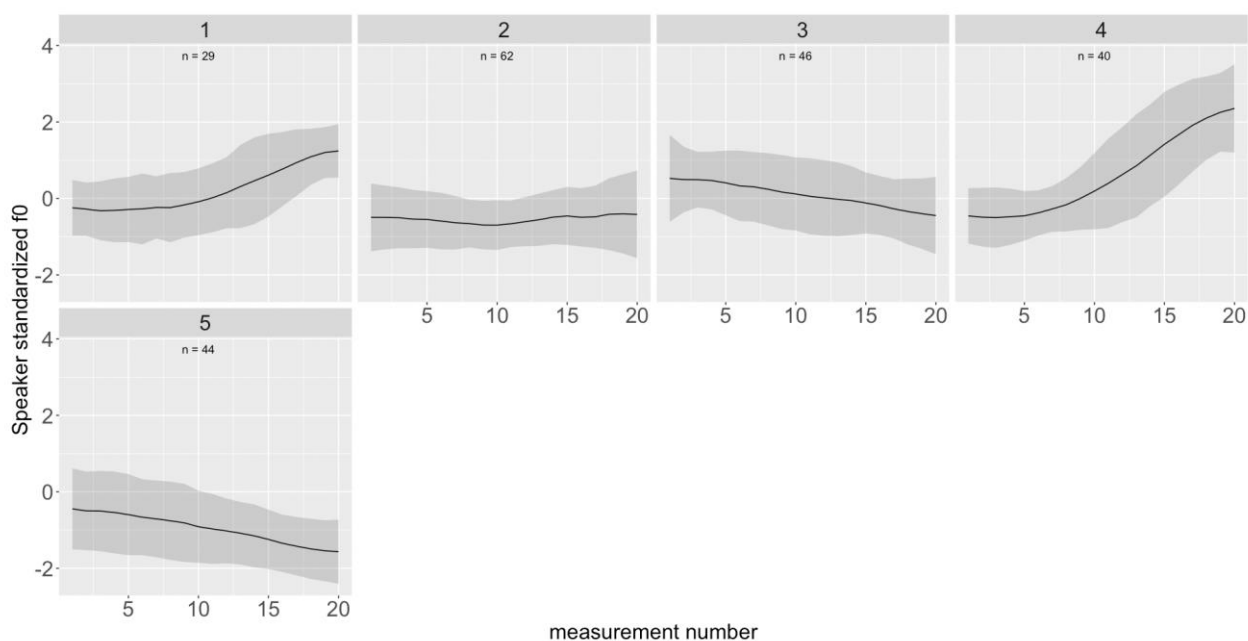


Рис. 1. Итоговые кластеры

Таким образом, в ходе кластерного анализа были получены следующие тональные типы:

- 1) Восходящий с меньшим диапазоном (1);
- 2) Восходящий с бóльшим диапазоном (4);
- 3) Ровный (2);
- 4) Нисходящий в более высоком регистре (3);
- 5) Нисходящий в более низком регистре (5).

Проанализировав автоматически сгенерированные textgrid-аннотаций с разметкой кластеров, мы отметили следующие тенденции:

- Ровный тон чаще всего встречается в утверждениях, наряду с нисходящим в более высоком регистре; в перечислениях; в специальных вопросах, но также и во многих других коммуникативных типах;

- Восходящий с бóльшим диапазоном встречается в перечислениях (часто — в первых синтагмах); в незавершенных высказываниях, наряду с восходящим с меньшим диапазоном; используется для выражения контраста (*Одна полностью// заполнена, вторая почти=наполнена.*); в специальных вопросах с частицей *a* (*А в чѐм// он еще раз одет?*);
- Нисходящий в более высоком регистре оформляет специальные вопросы (*На чѐм\ едет?*) и утверждения;
- Нисходящий в более низком регистре встречается в завершенных утверждениях, заканчивающих абзац; в конечной синтагме при перечислениях (необязательно завершенных); в уточняющих вопросах (*Это цветной\ фильм, да?*).

Наиболее интересными нам показались реализации тональных акцентов в перечислениях (см. рис. 2). Как можно видеть на рисунке, во всех синтагмах реализован один и тот же тональный паттерн — восхождение тона, однако различается тайминг: в первой синтагме восхождение начинается с предударного слога (4-ый тип акцента), во втором на предударном и ударном слогах наблюдается ровный тон с небольшим повышением и повышение происходит на заударных слогах (2-ой тип акцента), в конечной синтагме на ударном слоге происходит понижение тона и повышение происходит также на заударных слогах (3-ий тип акцента). Таким образом, мы предполагаем, что выбор тонального паттерна может быть обусловлен порядком синтагмы в перечислении.

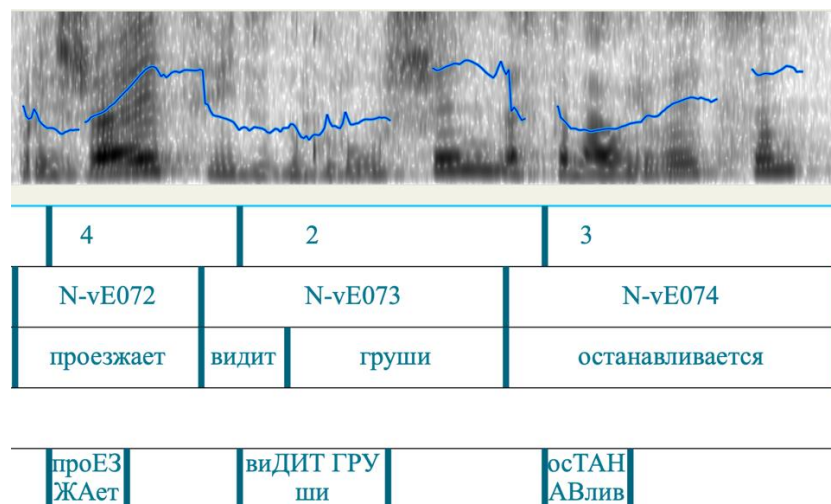


Рис. 2. Дистрибуция разных типов тонов в перечислениях (Динамическая спектрограмма с разметкой на ЭДЕ, слова и акценты (ударный и предударный слоги) сочетания синтагм в программе PRAAT (Проезжает//, видит груши=, останавливается\))

Кроме того, можно заметить, что во всех случаях движение тона не соответствует «эталонному» контуру интонации незавершенности — ИК-3 [14], т.к. основной подъем тона (как и пик ЧОТ) происходит на заударной части. Данное явление уже отмечалось в работе [13]; авторы отмечают, что такое тональное оформление незавершенных высказываний свойственно молодежной речи, что подтверждается нашими данными: возраст всех носителей — до 30 лет.

Также хотим отметить большую распространенность ровных тонов (которые, строго говоря, соответствуют динамическим акцентам), которые наблюдается практически во всех типах высказываний — мы предполагаем, что это также особенность речи молодежи.

Данный список тональных типов не претендует на полноту и является начальным этапом применения кластерного подхода к речевым данным на русском языке. В дальнейшем предполагается расширить базу анализируемых контуров и произвести также исследование «от функции к форме» - то есть отдельных коммуникативных категорий, в частности, перечислений: таким образом, можно будет уточнить дистрибуцию тональных паттернов. Кроме того, планируется провести экспериментальную проверку релевантности найденных категорий.

Список использованной литературы

1. *Кодзасов С. В., Кривнова О. Ф.* Общая фонетика. — Rossiiskii gos. gumanitarnyi universitet, 2001.
2. *Оде С.* Заметки о понятии тональный акцент на примере русского языка // Проблемы фонетики / под ред. Р. Ф. Касаткиной. М.: Наука. — 2007. — С. 237–249.
3. *Kaland C.* Contour clustering: A field-data-driven approach for documenting and analysing prototypical f0 contours // Journal of the International Phonetic Association. — 2021. — Pp. 1–30.
4. *Кибрик А. А.* Русский мультимедийный дискурс. Часть II. Разработка корпуса и направления исследований // Психологический журнал. — 2018, № 39(2). — С. 79–90. <https://multidiscourse.ru/corpus/>
5. *Кодзасов С. В.* Фазовая символика тона // Логический анализ языка. Семантика начала и конца / под ред. Н. Д. Арутюновой. — М.: Индрик, 2002.
6. *Николаева Т. М.* Акцентно-просодические средства выражения категории определенности-неопределенности // Категория определенности-неопределенности в славянских и балканских языках. — 1979. — С. 119–174.
7. *Nolan F.* Intonational equivalence: An experimental evaluation of pitch scales. Proceedings of the 15th international congress of phonetic sciences. — 2003, vol. 39. — Pp. 771–774.
8. *R Core Team.* (2020). R: The R project for statistical computing. Retrieved July 11, 2019, from <https://www.r-project.org/>
9. *Tran Dat & Michael Wagner.* 2002. Fuzzy C-means clustering-based speaker verification // Nikhil R. Pal & Michio Sugeno (eds). Advances in soft computing: AFSS 2002. — Berlin & Heidelberg: Springer, 2002. — Vol. 2275. — Pp. 318–324.

10. *Seeliger H., Kaland C.* Boundary tones in German wh-questions and wh-exclamatives—a cluster-based approach // Proceedings of the 11th International Conference on Speech Prosody. — 2022.
11. *Calhoun S., Schweitzer A.* Can intonation contours be lexicalised? Implications for discourse meanings // Prosody and meaning. — 2012. — Pp. 271–328.
12. *Charrad M. et al.* NbClust: an R package for determining the relevant number of clusters in a data set // Journal of statistical software. — 2014, vol. 61. — Pp. 1–36.
13. *Вольская Н. Б., Скрелин П. А.* Проблема произносительной нормы в актерской речи // Речевое творчество актера: данность и предчувствие. — 2017. — С. 87–93.
14. *Брызгунова Е. А.* Звуки и интонация русской речи. — 5-е изд. — 1983.

СРАВНЕНИЕ МЕЛОДИЧЕСКИХ КОНТУРОВ ОДНОСЛОЖНЫХ СИНТАГМ В РУССКОМ И АНГЛИЙСКОМ ЯЗЫКАХ

Павел Андреевич Холявин

(Санкт-Петербургский государственный университет)

Наличие интонации, как было подчёркнуто ещё Ч. Хоккеттом [5], является одним из универсальных свойств языка. Большой интерес для науки представляют сравнительные и типологические исследования интонационных систем разных языков. Данное исследование принадлежит к числу тех, в которых сравнивается интонация русского языка (а именно современного русского литературного языка) и интонация британского варианта английского языка.

Минимальной линейной единицей, на которой реализуется интонация, является синтагма [1]. В ряде исследований предлагается членить синтагму на определённые составные части, в частности, интонационный центр или ядро, несущий основное ударение в синтагме.

Интонационная система русского языка была неоднократно описана исследователями. Одним из наиболее известных описаний является система, разработанная Е. А. Брызгуновой [2]. Согласно этой системе, в русском языке существует всего семь интонационных конструкций (ИК). Основным (но не единственным) критерием для различения интонационных конструкций является характер мелодического движения на интонационном центре. Н. Б. Вольской была разработана система, расширяющая и уточняющая описание Брызгуновой [8]. В этой системе всего выделяется 13 моделей (у ряда моделей выделяются также варианты).

Интонация британского варианта английского языка также неоднократно подвергалась описаниям. Одно из самых известных описаний принадлежит Дж. Д. О'Коннору и Г. Ф. Арнольду [7]. В рамках этого описания выделяется семь мелодических движений, которые могут присутствовать внутри интонационного центра (т. н. «ядерных тонов»). Сочетаясь с другими структурными составляющими, эти семь ядерных тонов могут образовывать десять тональных групп.

Несложно заметить, что во всех упомянутых описаниях особенное внимание уделяется движению в рамках интонационного центра. Ниже приведена таблица, в которой показано соответствие некоторых интонационных типов русского и английского языков на основании схожести движения в интонационном центре.

Соответствие русских и английских интонационных типов

ИК русского языка (по Е. А. Брызгуновой)	Ядерные тоны английского языка	Характер движения
ИК-1	Low Fall	Нисходящее
ИК-2	High Fall	Нисходящее с более высоким началом
ИК-4	Low Rise	Восходящее
ИК-4	Fall-Rise	Нисходяще-восходящее

ИК-4 появилась в таблице дважды, как соответствие восходящему ядерному тону Low Rise и нисходяще-восходящему движению Fall-Rise. Интонационная конструкция № 4 обыкновенно реализуется как восходящее движение, которому, однако, может предшествовать некоторое понижение. В английском языке же эти два движения чётко противопоставлены.

Согласно принципу замены, сформулированному А. М. Пешковским, интонационное выражение синтаксического значения тем слабее, чем сильнее оно выражено грамматическими средствами [3]. Таким образом, роль интонации должна возрастать при сокращении длины высказывания. Предельным случаем такого сокращения будут высказывания, состоящие из одной синтагмы, которая, в свою очередь, состоит лишь из одного односложного слова. Этот слог автоматически будет становиться ядром синтагмы. Предполагается, что на материале таких высказываний можно будет изучить системы элементарных мелодических движений языков и исследовать особенности их функционирования.

Материалом для данного исследования послужили корпуса диалогической спонтанной речи на русском и английском языках. Источником русской речи стал корпус SibLing [6], записанный на кафедре фонетики и методики преподавания иностранных языков СПбГУ. Корпус снабжён экспертной просодической разметкой, в которую входит расстановка границ между синтагмами в расшифровках текстов, определение места интонационного центра в каждой синтагме и реализованной интонационной модели по системе Н. Б. Вольской. Кроме того, файлы разметки корпуса содержат временные границы всех высказываний, порождённых дикторами. Для исследования была взята часть корпуса, содержащая диалоги, в ходе которых участники совместно выполняли задание на поиск соответствий на картинках (именно эта часть сопровождается просодической разметкой).

Источником английской речи стал корпус LUCID (London UCL Clear Speech in Interaction Database), записанный в Университетском колледже Лондона [4]. Для исследования была взята часть корпуса, содержащая диалоги между носителями южнобританского варианта английского языка (20 мужчин и 20 женщин от 19 до 29 лет), совместно выполнявшими задание на поиск различий на картинках. Корпус не снабжён какой-либо просодической аннотацией, однако файлы его разметки содержат информацию о временных границах высказываний и слов, их составляющих.

Оба корпуса были автоматически проанализированы, и из них были выделены высказывания, содержащие одно слово: «да» для русского языка и «yes» для английского. Таким образом, создаётся банк высказываний, имеющих схожую форму (состоящих из одного слога) и схожую функцию (подтверждение слов собеседника). Все полученные таким образом фрагменты были обработаны инструментом для вычисления частоты основного тона REAPER¹. Полученные значения были использованы для построения мелодического контура каждого высказывания.

Анализ формы каждого контура проводился по следующему алгоритму:

1. Значения частоты основного тона (F_0) переводились из герцев в полутона;
2. Применялся алгоритм сглаживания;
3. В контуре выбиралось три ключевых точки, соответствующих значению F_0 в начале и в конце звонкого участка синтагмы, а также промежуточному значению, которое выбиралось как локальный минимум или максимум либо, при отсутствии такового (т. е. если движение было полностью монотонным), значение в серединной точке звонкого участка.

Далее были построены графики, на которых каждое высказывание отмечалось точкой, абсцисса которой определялась перепадом между первой и второй ключевыми точками, а ордината — перепадом между второй и третьей. Таким образом, восходящие тоны будут находиться в первой четверти графика, нисходящие — в третьей, нисходяще-восходящие — о второй, восходяще-нисходящие — в четвёртой. График для русских высказываний приведён на рис. 1, для английских высказываний — на рис. 2.

¹ <https://github.com/google/REAPER>

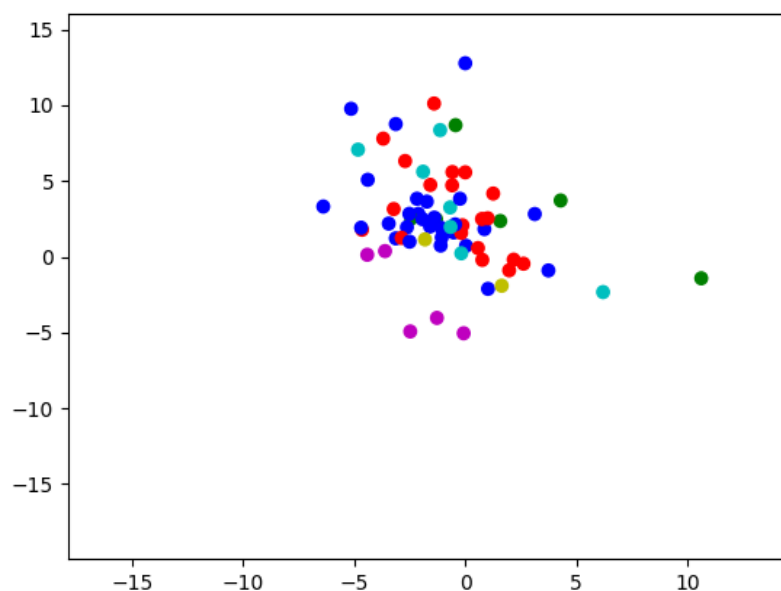


Рис. 1. Зависимость размаха второго перепада ЧОТ от первого для русских высказываний (полутона относительно среднедикторской ЧОТ)

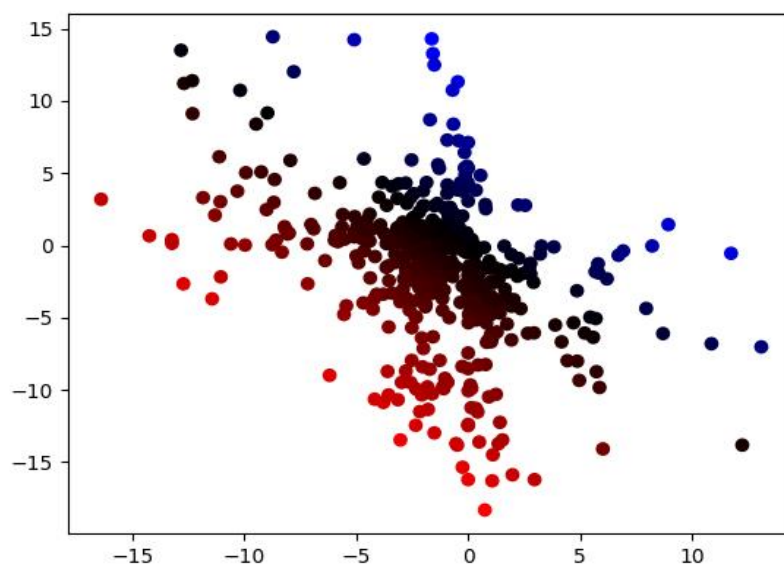


Рис. 2. Зависимость размаха второго перепада ЧОТ от первого для английских высказываний (полутона относительно среднедикторской ЧОТ)

Несложно заметить, что финальное движение гораздо чаще является нисходящим в английском материале, чем в русском. Это подтверждается слуховым анализом, который показывает преобладание в русском материале восходящих движений.

Поскольку данные для русского языка сопровождалась просодической разметкой, была определена частотность использования интонационных моделей. Данные приведены в таблице 2.

Встречаемость интонационных моделей в высказываниях, состоящих из слова «да»

Модель	ИК	Сколько раз встретилась
13	4	28
12	6	22
11	3	7
02	2	6
07	3	6
06	(1)	3

Можно наблюдать, что чаще всего встречаются модели, реализующиеся восходящими контурами и выражающие незавершённость — это модели 11, 12 и 13, соответствующие ИК-3, ИК-4 и ИК-6 в системе Е. А. Брызгуновой. Таким образом, результаты анализа F_0 подкрепляются данными экспертной просодической разметки.

Итак, исследование показало, что в высказываниях дикторов-носителей русского языка преобладали восходящие тоны (модели 11, 12 и 13 по системе Н. Б. Вольской), тогда как в высказываниях дикторов-британцев преобладали скорее нисходящие тоны. Известно, что как в английском, так и в русском языке нисходящие тоны служат для передачи значения завершённости. Отход от использования нисходящих тонов, более заметный в русском материале, чем в английском, может судить о разных дополнительных значениях, которые несут в себе восходящие тоны. Так, в английских утверждениях низкий восходящий тон может использоваться для того, чтобы показать собеседнику, что говорящий ещё не сформировал мнение относительно обсуждаемой темы и приглашает собеседника продолжить диалог. Вероятно, аналогичное значение можно было бы выделить и в русском языке. Однако в английском восходящий тон в утверждениях может нести и негативные коннотации (критика собеседника), чем, вероятно, можно было бы объяснить сниженную частотность его употребления.

Список использованной литературы

1. Светозарова Н. Д. Интонационная система русского языка. — Л.: Изд-во Лен. ун-та, 1982.
2. Брызгунова Е. А. Звуки и интонация русской речи. — 5-е изд. — М.: Русский язык, 1983.
3. Пешковский А. М. Интонация и грамматика // Избранные труды. — М.: Учпедгиз, 1959.

4. *Baker R., Hazan V.* LUCID: a corpus of spontaneous and read clear speech in British English // DiSS-LPSS Joint Workshop 2010, 2010.
5. *Greenberg J. H.* Universals of language // 1963.
6. *Kachkovskaia T. et al.* SibLing Corpus of Russian Dialogue Speech Designed for Research on Speech Entrainment // Proceeding of LREC, Marseille, 2020. — Pp. 6556–6561.
7. *O'Connor J. D., Arnold G. F.* Intonation of colloquial English. — London: Longman, 1973.
8. *Volskaya N., Kachkovskaia T.* Prosodic annotation in the new corpus of Russian spontaneous speech CoRuSS // Proceedings of Speech Prosody 2016. — Boston, 2016. — Pp. 917–921.

THE SECOND FORMANT DYNAMICS IN RUSSIAN AND CHINESE SYLLABLES

Guangqi Zhou

Svetlana O. Tananaiko

(Saint-Petersburg State University)

It is well known that the very presence of complex coarticulation processes is a universal phenomenon due to the structure of the human articulatory system. At the same time, however, the features of the coarticulation processes are specific for each language [3].

The objective of the study was to analyze the dynamics of the second formant in Russian and Chinese syllables of the CV structure with palatalized and palatal consonants. As it is known from acoustic phonetics [2], the second formant is associated with a vowel row, and since the presence of consonants with the main or additional palatal articulation before vowels inevitably shifts the vowel row in the direction of [i]- or [j]-shaped articulation (especially at the beginning of a vowel), the dynamics of the second formant in such syllables will be the main indicator of the consonant-to-vowel transition specificity in the syllables of this type. Russian syllables like “palatalized consonant + vowel” are frequent and typical, since the opposition of non-palatalized and palatalized consonants is one of the system-forming oppositions of Russian consonantism, and non-palatalized and palatalized consonants form parallel phonemic series, which include almost all consonants of the Russian language. In the theoretical descriptions of Chinese consonantism, there is no consensus on the phonological interpretation of palatalized and palatal sounds, opinions differ greatly, but the very existence of these sounds is indisputable. Articulatory descriptions of these sounds always record the process of palatalization, expressed in the presence of additional palatal articulation. It appears that if the sound is not followed by the front upper vowel [i] or [y], then the consonant is followed by the palatal glide [j] or [ɥ]. Therefore, syllables denoted in pinyin as beginning with *ji-*, *qi-*, *xi-*, *ju*, *qu-*, *xu-*, followed by a vowel, are described phonetically as beginning with [ts^j], [ts^{hj}], [s^j], [ts^ɥ], [ts^{hɥ}], [s^ɥ] [1].

In the course of the experiment, the dynamics of the second formant in Russian and Chinese syllables with a relatively similar sound were compared. The syllables contained palatalized labial and anterolingual consonants followed by back vowels [a], [o], [u], as well as a palatal approximant [j] in combination with the same vowels. Posterior consonants were not studied, since they are impossible in Chinese in this context. The speakers were 4 Russians (two men, two women, 18–20 years old), native speakers of Russian standard pronunciation, and 4 Chinese speakers (also two men, two women, 18–20

years old), native speakers of Chinese standard pronunciation. The speakers read words that included syllables with the studied sounds in similar (in Russian and Chinese) phonetic contexts. In Russian, words were selected where the studied syllables were stressed, in Chinese, words where the studied syllables were strong forms, and in both languages, the speakers were asked to read the words as clearly and legibly as possible. In total, 88 Russian and 88 Chinese words were read, in which palatalized consonants were before non-front vowels. The next stage of the work was the FII dynamics study in Russian and Chinese syllables. To do this, the duration of the [i]-shaped transition at the beginning of the second formant in each syllable was measured, its frequency difference (drop) was calculated, and then the frequency difference was calculated per 100 ms, which allowed comparing the steepness of the formant drops for each vowel in Russian and Chinese (see Fig. 1, where the rectangle is the zone [i]-shaped transition), after what the average FII drop per 100 ms was calculated for different groups of consonants.

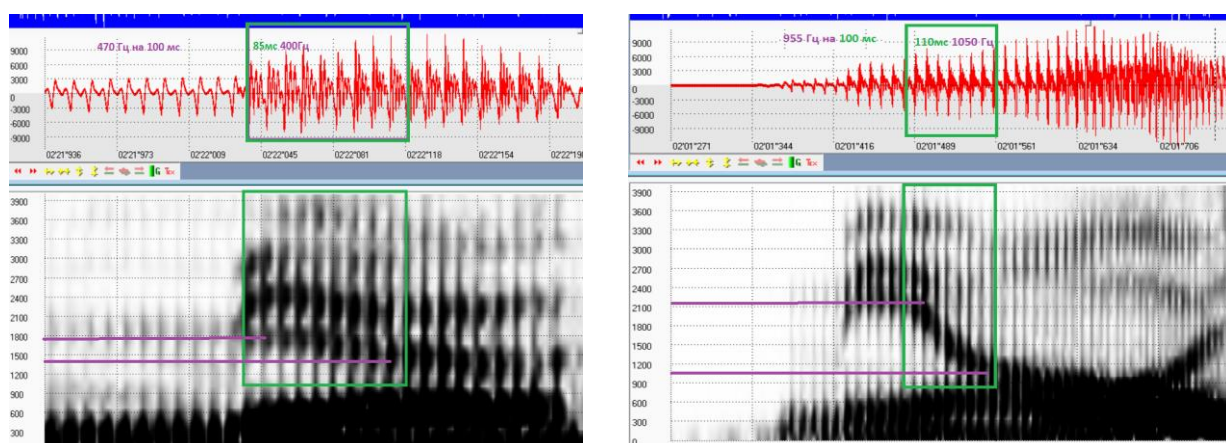


Fig. 1. The oscillogram and the spectrogram of the syllable [m'a] in the Russian word *мясо* [m'asa] (on the left) and the Chinese word *miao xiao* (on the right)

According to the results of statistical analysis, the following data were obtained.

In both languages, the value of the FII difference (see Fig. 2) for the sound [j] exceeds all other values, and at the same time is almost the same in both languages. This is easily explained: it is during the articulation of the palatal approximant [j] that the middle of the tongue rises as high as possible, which leads to a strong difference between the second formant of [j] (about 2500 Hz) and the second formant of non-front vowels. However, a more detailed analysis shows that the drop after [j] is different for different vowels in both languages (see Fig. 3): for [a] and [o], it is less in Russian than in Chinese, and for [u], on the contrary, in Russian the value is greater. The excess of Russian

drop for [u] is such that it equalizes the drop average values for both languages. For labial consonants (see Fig. 2) the difference is greater in Chinese than in Russian, and this trend persists for different vowels (see Fig. 3). For the anterolingual in general (see Fig. 2) in Russian the difference is greater than in Chinese, but this excess is achieved due to (see Fig. 3) FII dynamics of labialized vowels [o] and [u]. For [a] the result is opposite, Russian has a larger FII difference, i.e. in Russian the FII value is lower than in Chinese, in other words, Russian [a] is pronounced as a more retracted vowel compared to Chinese in these consonant contexts.

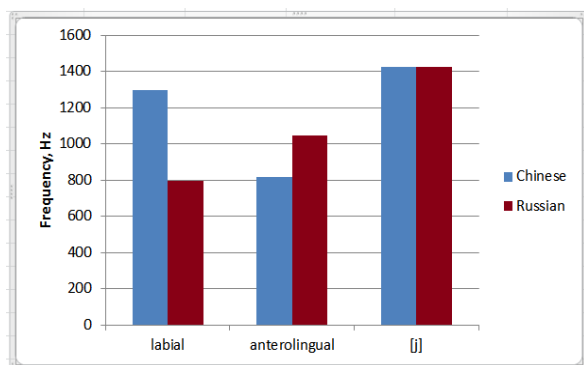


Fig. 2. FII dynamics after different types of consonants, the data is averaged over all vowels

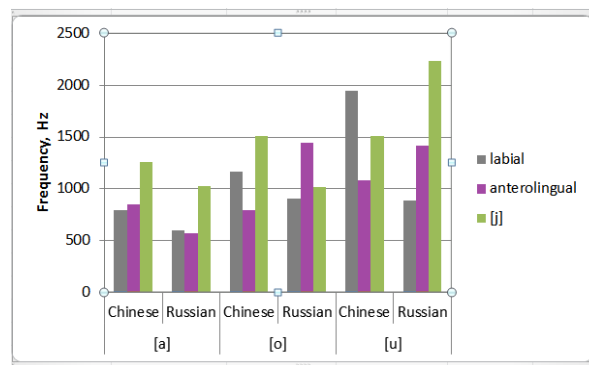


Fig. 3. FII dynamics after different types of consonants before different vowels

The dependence of the drop on the vowel is obvious (see Fig. 4): the difference is the greater the more a vowel is retracted: in both languages it is least for [a], more for [o] and maximum for [u], which is explained by the articulatory and, consequently, acoustic characteristics of the corresponding sounds: among the three studied vowels, the vowel [u] has the minimal second formant, and for [a] it is maximal. At the same time, for all vowels, the drop in Chinese slightly exceeds the drop in Russian, and this excess is maximum for [a] and minimum for [u].

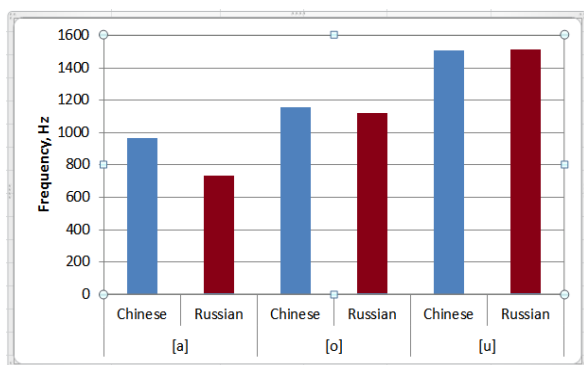


Fig. 4. FII dynamics for different vowels, the data is averaged over all consonants

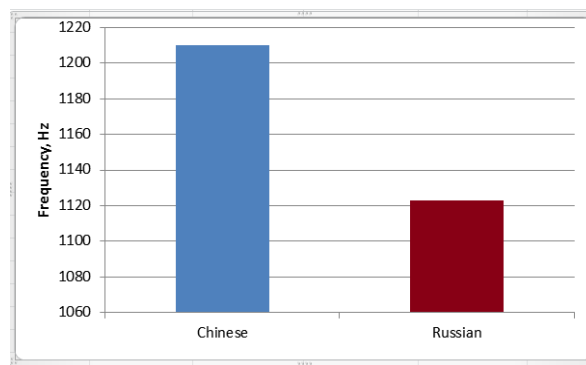


Fig. 5. FII dynamics in Chinese and Russian, averaged data

The average values of the FII difference (see Fig. 5), i.e. the values averaged over all groups of consonants and over all vowels are somewhat larger in Chinese than in Russian.

These data allow us to conclude that the syllabic structure of the Chinese language and the integrity of the syllable as a meaningful unit does not necessarily imply a greater articulatory and, consequently, acoustic interdependence of sounds compared to the Russian language.

Список использованной литературы

1. *Ladefoged Peter, Wu Zongji. Places of Articulation: An Investigation of Pekingese Fricatives // Journal of Phonetics. — 1984, vol. 12. — Pp. 267–278.*
2. *Бондарко Л. В. Фонетика современного русского языка. — СПб., 1999.*
3. *Кодзасов С. В., Кривнова О. Ф. Общая фонетика. — М., 2001.*

ABSTRACTS OF PAPERS PRESENTED AT WORKSHOP “ANALYSIS OF THE RUSSIAN COLLOQUIAL SPEECH 2023” (IN ENGLISH)

FINAL VOCALIZATION IN ORAL DISCOURSE AS A SPEECH MANIPULATION TECHNIQUE (FALSE EXPECTATION EFFECT)

Natalia V. Bogdanova-Beglarian

The article attempts to test the hypothesis about the ability of vocalizations as a very common hesitation phenomenon in oral discourse to act as a method of covert manipulation. Saying *e-e* or *m-m*, the speaker seems to warn the interlocutor that he has not finished his remark yet, but is simply thinking about its continuation. It was assumed that such a signal would be correctly perceived by the interlocutor, who would not seize the speaking queue, but would wait until the speaker passed this queue to him. However, the analysis of corpus material did not confirm this hypothesis. Three types of situations were found in the considered contexts: (1) the interlocutor does not respond to the “warning” vocalization and enters into the conversation, interrupting the speaker; (2) the interlocutor perceives the signal to wait, but the speaker refuses to continue for some reason - and then the interlocutor enters into the conversation without interruption; (3) the interlocutor perceives vocalization as a request for help and prompts the speaker to the right word according to the “backchannel” principle. The expected function of vocalization as a manipulative unit turned out to be unrealized in all cases. The study was carried out on the material of the oral subcorpus (OS) of the National Corpus of the Russian Language and the corpus of the Russian language of everyday communication “One Speech Day” (ORD).

THE USAGE OF MODERN HOPFIELD NETWORKS FOR SPEECH RECOGNITION NEURAL NETWORK ALGORITHMS TRAINING ON UNLABELED DATA

Daniil V. Grebenkin

This paper considers the possibility of replacing the self-attention with dense associative memory layers in the wav2vec2 neural network automatic speech recognition (ASR) algorithm. The work examines the hypothesis that

modern Hopfield networks concept is more suitable for restoration of missing fragments of the audio signal task and speech-to-text task, than multi-head attention because the ASR model with trained associative memory layers contains the images of phoneme representations from natural language which can be recovered from any type of input audio. We created two ASR wav2vec2 models to check our assumption: with self-attention layers (classic model) and with dense associative layers (new architecture model). Experiments on several speech augmented and non-augmented corpora have shown that the model with the new architecture allows to improve the quality of speech recognition and it can be used for pretraining the models on a big amount of audiodata. The UMAP-projections and Silhouette coefficient were used to interpret the results.

A COMPARISON OF VOWEL FORMANT TRANSITIONS IN DIFFERENT LANGUAGES

*Vera V. Evdokimova
Maria R. Maksimova*

This paper compares formant transitions in English, French and German monophthongs after labial and lingual (namely, tip, front and back) consonants. The monophthongs were extracted from words, sentences and texts read by female speakers. The formant values were measured in 9 points within each vowel. The line graphs representing formant transitions were plotted. The derivatives of the resulting functions at 9 points were calculated to compare the slopes of different formant trajectories. The formant values were converted from Hertz to Bark in order to evaluate the variability of the formant structure. The results indicated that the formant transitions in English and German vowels were similar after tip and front consonants. In labial consonantal context, the dynamics of formant transitions did not show any clear tendencies. After back lingual consonants, the formant transitions were similar for all the three languages. The region in which the greatest slope occurred was the same for French and German monophthongs, but different for English vowels. The variability of formant structure was lower in the back monophthongs compared to the front vowels in English and German data. However, in French monophthongs, the transition from one critical bandwidth to another became more frequent with backness.

UNIVERSAL AND UNIQUE FEATURES IN THE INTONATION OF FRENCH INTERROGATIVE UTTERANCES

Olga V. Kamrysh

The paper presents a study of the universal and specific intonational features in Russian and French utterances, which appears important both for solving applied tasks of automatic speech synthesis and for identifying the linguodidactic foundations of teaching a foreign language. The material for the study included audio tracks of contemporary French films, freely available podcasts, as well as audio recordings of authentic study guides. The results of this experimental study suggest that the traditionally recommended model with the descending tone movement from the beginning to the end of the utterance is not the only possible one in the modern French language. When analyzing special questions with different syntactic organizations, it was demonstrated there were also other ways of melodic arrangement, including the use of even and ascending tones. We can also testify to the universality of the use of melody to convey emphatic meanings. By way of contrast, the falling-rising contour, similar in melodic arrangement, has different meanings in Russian and French. The experimental data obtained indicate that this contour appears to be one of the ways of expressing some emotional meanings in French, unlike in Russian, where the rising-falling tone movement is characteristic of emotionally neutral utterances.

COMPARATIVE ANALYSIS OF SPEAKER'S RESPIRATORY ACTIVITY AT REST AND WHEN READING PROSAIC AND POETIC TEXTS (UPON RESULTS FROM ONLINE MRI)

*Galina E. Kedrova
Maria V. Volkova
Nikolay V. Anisimov*

The paper presents preliminary results of the MRI investigation of respiratory strategies of Russian speakers when reading a poem and a prosaic text (short story), in comparison to their breathing in silence. MRI protocols optimized for online visualization of lung function were applied to register breathing behavior of 4 native Russian speakers. To obtain a numerical value for the total area of active segments of lungs in each reading mode, compared to the

one in rest, special processing of collected MRI images has been carried out. For each respiratory mode, the minimum and maximum values of the total area of active segments of the lungs, as well as median values and dispersion were calculated. The paper presents digitalized data of the breathing activity for one speaker. These results show an approximately 1.6-fold increase in the total area of active segments of the right and left lungs during reading compared with similar data for silence. The results also demonstrated constancy of increased lung volume throughout the process of phonation. Verification of the data with Mann–Whitney U test showed systemic differences between respiratory processes in speech production and breathing in silence.

MELODIC PATTERN OF FINAL AND NON-FINAL INTONATION GROUPS IN RUSSIAN SPEECH OF BRASILIANS

*Uliana E. Kochetkova
Tatiana V. Kachkovskaia
Arina A. Soboleva*

This paper presents the results of an acoustic analysis of the intonation of the Russian speech of Brazilians. Intonational interference in the speech of Brazilians in the melodic pattern of final and non-final intonation group was studied. It is shown how the melodic pattern of final intonation groups in Brazilian Portuguese influences the melodic pattern in the Russian speech of Brazilians. Results demonstrate high rank of interference in final intonation groups, for which the Brazilians tend to use melodic patterns that are normally used in contrastive Russian sentences. The analysis of melodic figure in non-final intonation groups was impeded due to the existence of different types of melodic patterns in interfered speech of Brazilians that resemble similar Russian patterns and thus need to be studied in more detail in order to show whether any specific difference exists.

ANALYSIS OF CHILDREN'S EMOTIONAL SPEECH

*Elena E. Lyakso
Olga V. Frolova
Alexandre S. Nikolaev*

The paper describes approaches to recording child's emotional speech ("Child Emotion Development Method" — CEDM), the results of perceptual

experiments and automatic classification of child's emotional states. The cross-cultural study of children speaking Russian and Tamil (India) languages was carried out. It was shown that native speakers recognize better emotional states by the speech of children who speak the same language. The correlation between the emotional state of children and the acoustic features of their emotional speech correctly recognized by Russian and Indian experts was revealed. The study of emotional state classification by human and automatically by the speech of typically developing children (TD), children with autism spectrum disorders (ASD) and Down syndrome (DS) was conducted. Listeners recognize better the state of discomfort and comfort by the speech of children with ASD and DS than by the speech of TD children. Automatic classification revealed that the emotional speech of children with TD is classified worse than speech of children with atypical development. The factors influencing auditors' recognition of the emotional state of children with TD and ASD were identified. The results of the study can be used for estimation of the emotional sphere in children with typical and atypical development.

CURRENT ISSUES OF AUTOMATIC SPEECH RECOGNITION: DETECTION AND ANALYSIS OF EXTRA-LINGUISTIC VOCALIZATIONS IN SPONTANEOUS SPOKEN SPEECH

*Anastasiia A. Povolotskaia
Alexey A. Karpov*

In this article the current state of automatic speech recognition (ASR) technology is reviewed. It is suggested that the improvement of algorithms and methods of detection and analysis of extralinguistic vocalizations (EV) in spontaneous spoken speech will not only improve the accuracy of speech recognition systems, but also the possibility of integrating these algorithms and methods in other areas of human activity will allow to extract more information from human's speech, specifically from non-verbal vocalizations, and recognize psycho-emotional and physical states of humans more accurately. A simplified basic method of speech signal processing for transcribing audio speech into text is presented, an original systematization of EV in Russian is given. Based on the analyzed literature and according to the compiled systematization of EV, a method of speech signal processing for identifying EV in spontaneous spoken speech is proposed and described.

ON THE NEURAL BASIS OF THE SCALES FOR ESTIMATING SHORT TIME INTERVALS

Daria N. Podvigina

Timing is an essential part of all sensory, motor and cognitive processes, although the neural basis of time perception is poorly understood. Many theories and models have been proposed to account for timing. Overall, these models can be divided into two broad classes: dedicated and intrinsic models. Dedicated models propose that the brain has a centralized set of circuits for timing that account for timing across modalities, tasks, and scales within the range of hundreds of milliseconds to seconds. Neurophysiological data strongly support intrinsic models, which propose that timing is an intrinsic computation of most neural circuits, and timing per se emerges from general properties of neurons, their temporal characteristic and the inherent dynamics of neural circuits [Paton & Buonomano, 2018]. Altogether, neurons form a “sensory scale” for time perception. According to psychophysical research, this scale is discrete with a limited number of gradations and can be adapted (rescaled) for any time range from tens of milliseconds to seconds.

ANALYSIS THROUGH SYNTHESIS: ESTABLISHING ACOUSTIC CUES OF IRONY BY MEANS OF AUDIO SIGNAL MODIFICATIONS

*Pavel A. Skrelin
Uliana E. Kochetkova
Vera V. Evdokimova
Daria D. Novoselova*

The method of modifications of acoustic characteristics is actively used to study various linguistic phenomena. It consists in transplanting certain prosodic parameters. We used it to test the individual and complex significance of the acoustic characteristics in the perception of of irony in speech. The experiments carried out included various modifications of the sound signal, including the transplantation of the melodic contour. At first the perceptual tests were carried out. A series of modifications included the change of the duration and intensity of the stressed vowel; the change of the melodic contour of the target fragment and the combination of changes. The obtained stimuli have been tested in perceptual experiments. The most important acoustic parameter responsible for ironic assessment is melodic changes (including combinations of intonational contour changes with other acoustic characteristics).

APPROXIMATION AND HEDGING IN LANGUAGE AND SPEECH

*Yanan Xiang,
Natalia V. Bogdanova-Beglarian*

The report is devoted to the analysis of the phenomena of approximation and hedging in the language and the specifics of the use of pragmatic markers—approximators in everyday Russian speech (*kak by, tipa, ili tam, v lode* etc.). Such markers show the speaker's uncertainty about what he is talking about or is using when the direct naming of an object, phenomenon or situation is redundant, inappropriate, or impossible. Markers—approximators are one of the varieties of hedging, i.e., the use of words (units) of “cautious” modality — indefinite pronouns or adverbs, as well as introductory words that mark the speaker's uncertainty (*neskol'ko* ‘some’, *koe-gde* ‘somewhere’, *vidimo* ‘probably’, *kazhets'a* ‘seem’ etc.). Both strategies — both approximation and hedging — are aimed at weakening the illocutionary force of the statement, they equally make the statement vague, partially relieve the speaker of responsibility for what was said and soften the categoricalness of his statements. The study was conducted on the material of the spoken subcorpus of Russian National Corpus. Its results can serve as material for works on colloquial studies and are also important in the practice of teaching Russian as a foreign language and in the practice of translating Russian literary texts into other languages.

AUTOMATIC CLUSTERING OF PITCH ACCENTS IN RUSSIAN

Elizaveta V. Fanenshtyl

We present data from a study investigating pitch accent realizations in Russian. We focus here on the variety of prosodic contours found in the data and propose to a cluster approach to classify them. It consists in dividing a set of tonal contours into homogeneous groups; it allows semi-automatic processing of a large amount of data and eliminates possible bias towards the object of study. We examined speech of 6 native speakers of Russian and performed a cluster analysis towards 408 prosodic contours, clustering algorithm is hierarchical agglomerative with a time-warping based distance measure. To determine optimal number of clusters we applied a variety of statistical tests using an R package. The results indicate 5 tone patterns: two falling contours with different tone registers, two rising ones with different tone ranges and a

flat contour. The paper also presents the variety of prosodic patterns found in enumeration.

COMPARING MELODIC CONTOURS IN ONE-SYLLABLE INTONATION PHRASES IN RUSSIAN AND ENGLISH

Pavel A. Kholiavin

The study compares the melodic contours in one-syllable intonation phrases in spontaneous dialogue speech in two languages: Russian and British English. The specific types of intonation phrases chosen for analysis were those expressing affirmation: “да” (‘yes’) for Russian and “yes” for English. The results show that Russian phrases are more likely to exhibit rising intonation; this can be attributed to the fact that, while in both languages rising tones can be employed to signal readiness to continue the dialogue, in English low rising tones can also convey a negative attitude (contradiction and resentment).

THE SECOND FORMANT DYNAMICS IN RUSSIAN AND CHINESE SYLLABLES

Zhou Guanqi

Svetlana O. Tananaiko

The objective of the study was to analyze the second formant dynamics in Russian and Chinese CV syllables with palatalized and palatal consonants. The material for the study was read speech of Russian and Chinese speakers: 4 Russians (two men, two women, 18-20 years old), native speakers of the Russian standard pronunciation, and 4 Chinese (also two men, two women, 18-20 years old), native speakers of the Chinese standard pronunciation. The speakers read words that included CV syllables with back vowels in similar (in Russian and Chinese) phonetic contexts. In Russian, words were selected where the studied syllables were stressed, in Chinese, words where the studied syllables were strong forms. In total, 88 Russian and 88 Chinese words were read. Then the FII dynamics in Russian and Chinese syllables was studied. The received data allow to conclude that the syllabic structure of Chinese language and the integrity of the syllable as a meaningful unit does not necessarily imply a greater articulatory and, consequently, acoustic interdependence of sounds compared to the Russian language.

Научное издание

АНАЛИЗ РАЗГОВОРНОЙ РУССКОЙ РЕЧИ
АР³ — 2023

Труды десятого междисциплинарного семинара

Подписано в печать 25.06.2023. Формат 60 × 84 ¹/₁₆.
Усл. печ. л. 5,6. Заказ № 14576.

ООО «Издательство Скифия-принт»
197198. С.-Петербург, ул. Б. Пушкарская, д. 10, лит. А. пом. 32-Н