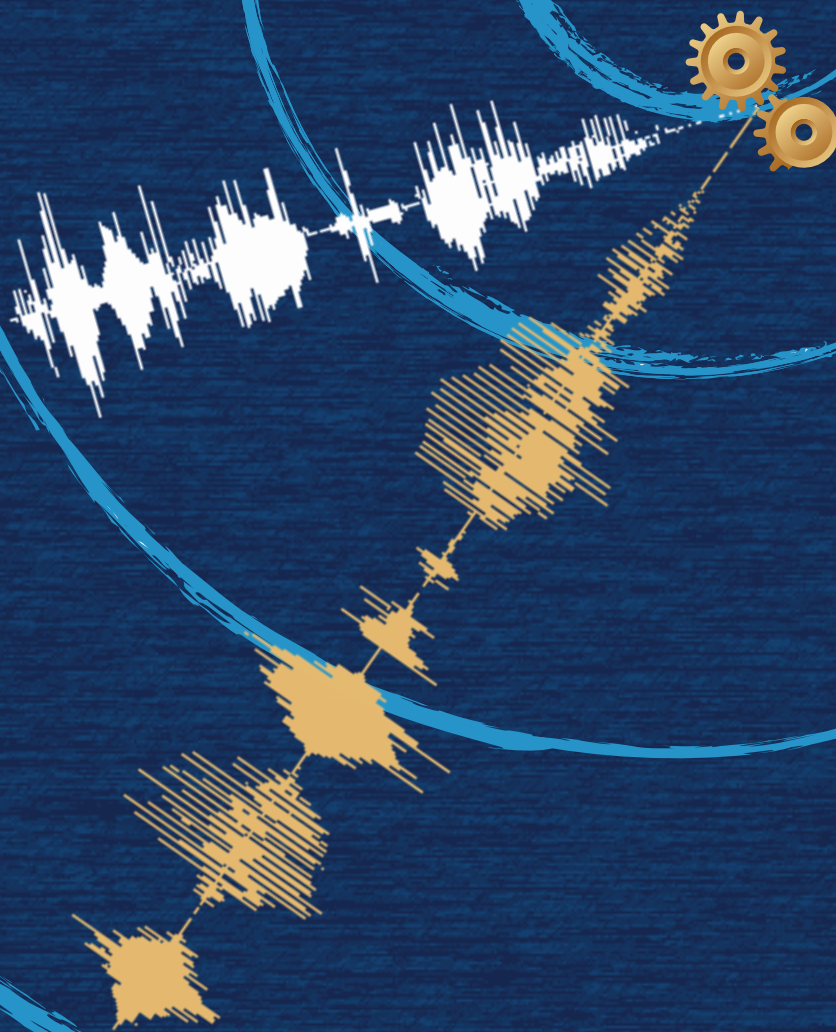


НАУКА  
ТЕХНОЛОГИИ  
ИНЖЕНЕРИ  
ЭКОНОМИКА  
И  
СОЦИАЛЬНЫЕ  
НАУКИ

# РЕЧЕВЫЕ ТЕХНОЛОГИИ



Speech technology

3-4

2021



# Речевые

## ТЕХНОЛОГИИ

3–4/2021

**Главный редактор:** *Харламов Александр Александрович*,  
доктор технических наук, [kharlamov@analyst.ru](mailto:kharlamov@analyst.ru)

**Состав редколлегии:**

Заместитель главного редактора: *Потапова Родмонга Кондратьевна*,  
доктор филологических наук, профессор, [rkpotapova@yandex.ru](mailto:rkpotapova@yandex.ru)

*Азаров Илья Сергеевич*, доктор технических наук, профессор,  
[azarov@bsuir.by](mailto:azarov@bsuir.by), Беларусь

*Голенков Владимир Васильевич*, доктор технических наук, профессор,  
[golen@bsuir.by](mailto:golen@bsuir.by), Беларусь

*Дмитриев Владимир Тимурович*, кандидат технических наук,  
[vol77@rambler.ru](mailto:vol77@rambler.ru)

*Железны Милош*, PhD,  
[zelezny@kky.zcu.cz](mailto:zelezny@kky.zcu.cz), Чехия

*Жигулёвцев Юрий Николаевич*, кандидат технических наук,  
[ynzh@mail.ru](mailto:ynzh@mail.ru)

*Карпов Алексей Анатольевич*, доктор технических наук,  
[karpov@iias.spb.ru](mailto:karpov@iias.spb.ru)

*Коренева Ольга*, PhD,  
[quovadis36@gmail.com](mailto:quovadis36@gmail.com), Испания

*Кудубаева Сауле Альжановна*, кандидат технических наук,  
[saule\\_58@mail.ru](mailto:saule_58@mail.ru), Казахстан

*Кушнир Алексей Михайлович*, кандидат психологических наук,  
[kushnir-narobr@yandex.ru](mailto:kushnir-narobr@yandex.ru)

*Кушнир Дмитрий Алексеевич*, кандидат технических наук,  
[kushdal@yandex.ru](mailto:kushdal@yandex.ru)

*Лобанов Борис Мефодьевич*, доктор технических наук,  
[lobanov@newman.bus-net.by](mailto:lobanov@newman.bus-net.by), Беларусь

*Ляксо Елена Евгеньевна*, доктор биологических наук,  
[lyakso@gmail.com](mailto:lyakso@gmail.com)

*Максимов Евгений Михайлович*, доктор технических наук,  
[maximovem@inbox.ru](mailto:maximovem@inbox.ru)

*Мещеряков Роман Валерьевич*, доктор технических наук, профессор,  
[trv@ieee.org](mailto:trv@ieee.org)

*Пильгун Мария Александровна*, доктор филологических наук,  
[pilgun@yandex.ru](mailto:pilgun@yandex.ru)

*Ромашкин Юрий Николаевич*, кандидат технических наук,  
[romayn@yandex.ru](mailto:romayn@yandex.ru)

*Ронжин Андрей Леонидович*, доктор технических наук,  
[ronzhin@iias.spb.ru](mailto:ronzhin@iias.spb.ru)

*Славкова Светлана*, PhD,  
[svetlana.slavkova@unibo.it](mailto:svetlana.slavkova@unibo.it), Италия

*Смирнов Иван Валентинович*, кандидат физико-математических наук,  
[ivs@isa.ru](mailto:ivs@isa.ru)

*Сулейманов Джавдет Шевкетович*, академик Академии наук Татарстана,  
профессор, [alsu\\_73@list.ru](mailto:alsu_73@list.ru)

*Чучупал Владимир Яковлевич*, кандидат технических наук,  
[v.chuchupal@gmail.com](mailto:v.chuchupal@gmail.com)

## Содержание

<i>Пантюхин Д.В.</i>	
<b>Нейронные сети синтеза речи голосовых помощников и поющих автоматов . . . . .</b>	<b>3</b>
<i>Харченко С.С., Крючков И.А., Мещеряков Р.В.</i>	
<b>Определение частоты основного тона речевого сигнала на основе индекса многомерной синхронизации . . . . .</b>	<b>17</b>
<i>Борисов В.В., Харламов А.А.</i>	
<b>Интеллектуальный анализ и моделирование развития ситуаций на основе сочетания нейросетевой обработки текстов, методов нечеткого когнитивного анализа и динамической кластеризации . . . . .</b>	<b>29</b>
<i>Сейедмилад Ранаей Сиадат</i>	
<b>Распознавание эмоций в персидской речи с помощью одномерной (1D) нейронной сети . . . . .</b>	<b>44</b>
<i>Пильгун М.А.</i>	
<b>Когнитивные исследования в режиме реального времени: нейросетевой подход . . . . .</b>	<b>57</b>
<i>Коренева О., Лимбах К., Штендер А., Гор М.</i>	
<b>Нейродидактика и преподавание иностранных языков и перевода в медиaprостранстве – на примере немецкого подкаста университета Испании . . . .</b>	<b>71</b>
<i>Ана Медина Регера</i>	
<b>Речевые технологии развития общения у детей с отсутствием речи . . . . .</b>	<b>84</b>
<i>Изабель Гальего Гальярдо</i>	
<b>Понятие «устойчивый туризм» в социальной, экономической, экологической и лингвистической сферах . . . . .</b>	<b>100</b>
<i>Милица Лазович, Тимо Алерс</i>	
<b>Обучение немецкому языку как иностранному в тандеме с помощью приложения Holoingo!: анализ коммуникации в тандеме в игровой социальной виртуальной реальности . . . . .</b>	<b>113</b>
<b>Памяти А.М. Кушнира . . . . .</b>	<b>135</b>
<b>Памяти Ю.Н. Ромашкина . . . . .</b>	<b>137</b>

### Редакция:

Редактор: *Татьяна Иванова*  
 Корректор: *Людмила Асанова*  
 Дизайн: *Анна Ладанюк*  
 Вёрстка: *Андрей Кинсбургский*

**Адрес редакции:** 109341, Москва, ул. Люблинская, д. 157, корп. 2.

**Тел.:** (495) 345 52 00

**Электронная почта:** narob@yandex.ru

**Сайт:** www.narodnoe.org

Подписано в печать 26.12.2021. Формат 60×90%. Бумага офсетная. Печать офсетная.

Печ. л. 14,25. Заказ № 22420. Издательский дом «Народное образование».

Отпечатано в типографии НИИ школьных технологий.

109341 Москва, ул. Люблинская, д. 157, корп. 2 Тел.: (495) 345 52 00/59 00.

# Нейронные сети синтеза речи ГОЛОСОВЫХ ПОМОЩНИКОВ И ПОЮЩИХ АВТОМАТОВ

*Пантюхин Д.В., старший преподаватель, Национальный исследовательский университет «Высшая школа экономики», Москва, Россия, старший преподаватель, МИРЭА — Российский технологический университет, Москва, Россия, dpantiukhin@hse.ru*

В работе представлен обзор публикаций последних лет по использованию нейросетевых методов для синтеза речи. Дается краткое введение в нейросетевые архитектуры для синтеза речи. Рассмотрены методы создания певческого голоса для голосовых помощников и поющих автоматов.

• нейронные сети • поющие автоматы • синтез голоса.

## 1. ПРЕДСТАВЛЕНИЕ ЗВУКА И РЕЧИ

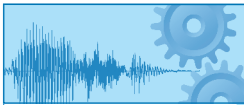
В компьютерах звук и речь для последующего анализа принято представлять в виде массивов чисел, обычно в двух видах:

- представление в виде звуковой волны — последовательности амплитуд записанного звука, выбранных с определенной частотой. Тогда речь представляется одномерным массивом чисел или набором таких одномерных массивов в случае многоканального звука.
- представление в виде спектрограмм различного вида — последовательности наборов частот, присутствующих в звуке в заданный период времени, выбранных с определенным шагом по частоте и по времени. Тогда речь представляется двумерным массивом чисел или набором таких массивов, которые можно интерпретировать как изображения и, следовательно, применять архитектуры для обработки изображений.

В спектральном представлении часто используют мел-спектрограммы, когда вместо частот используют мел-частоты — пересчитанные единицы для лучшего соответствия восприятию звука человеком. Также вместо самих частот используются коэффициенты набора фильтров, настроенных на разные частоты, среди них выделяют мел-кепстральные коэффициенты<sup>1</sup> и вейвлет-коэффициенты<sup>2</sup>. В любом случае

<sup>1</sup> Подробное объяснение и визуализацию см., например, в <https://habr.com/ru/post/140828/>

<sup>2</sup> Подробное объяснение и визуализацию см., например, в <https://habr.com/ru/post/449646/>



набор таких коэффициентов, изменяющихся во времени, можно представить двумерным массивом.

Амплитуда звука сильно подвержена искажениям, сильнее, чем частоты звука, поэтому методы анализа на основе представления звуковой волны до последнего времени были развиты не очень сильно. Но последние годы такие методы привлекают гораздо больше внимания, поскольку появились архитектуры и методы, которые достаточно устойчивы к подобным искажениям. Для задачи синтеза речи конечный продукт — это именно звуковая волна, которая будет воспроизводиться на устройстве. В случае со спектральным представлением потребуется перевод его в звуковую волну.

## 2. ЗАДАЧИ РАСПОЗНАВАНИЯ И АНАЛИЗА РЕЧИ

Задачи синтеза, создания речи неразрывно связаны с задачами анализа и распознавания. Многие архитектуры для синтеза речи содержат в себе части, обученные на задачах распознавания речи. Отметим важные направления в обработке речи<sup>3</sup>.

**A) Распознавание речи.** Это наиболее крупная и наиболее изученная категория задач. Многие другие задачи используют подходы, предложенные для распознавания речи.

Первая подзадача — это собственно распознавание речи — перевод звука в текст, соответствующий этому звуку. Существует более 60 различных наборов данных и более 100 тестов для них, содержащих записи звука и соответствующие им тексты, собранные в различных условиях, содержащие тысячи часов записанного звука на разных языках, с одним или несколькими дикторами, записи речи в толпе с шумами и искажениями<sup>4</sup>.

Другая подзадача связана с распознаванием речи не из звука ее, а по движениям губ, записанных на видео<sup>5</sup>. Решается и похожая задача, когда движения губ переводят не в текст, а сразу в синтезированный звук<sup>6</sup>. Возможны и другие визуальные источники для распознавания речи, например движения рук, пальцев и др. частей тела.

Распознавание речи, записанной на относительно большом расстоянии от микрофона, имеет затруднения из-за наличия шумов, реверберации и других искажений. Вариант решения — использование микрофонной решетки, совокупности нескольких микрофонов вместо одного<sup>7</sup>. Это накладывает требования и на архитектуры для распознавания речи.

<sup>3</sup> Согласно категориям <https://paperswithcode.com/area/speech>

<sup>4</sup> См. <https://paperswithcode.com/datasets?task=speech-recognition>

<sup>5</sup> См. <https://paperswithcode.com/task/visual-speech-recognition>

<sup>6</sup> См. <https://paperswithcode.com/task/lip-to-speech-synthesis>

<sup>7</sup> См. <https://paperswithcode.com/task/distant-speech-recognition>

Другая проблема при распознавании речи — это сильно выраженный акцент речи не носителей языка. Это затрудняет распознавание речи методами, которые используют только один язык. Вариант решения — использовать архитектуры, которые обучаются распознаванию речи сразу на нескольких языках, например английском и китайском<sup>8</sup>.

В некоторых применениях задачу полного распознавания речи редуцируют до распознавания только отдельных элементов ее: ключевых слов<sup>9</sup> (например, для чат-ботов служб технической поддержки пользователей), именованных сущностей (например, для опросов или перепроверок анкет пользователей), заранее определенных команд (например, для управления роботами). В таких случаях применяемые архитектуры могут быть существенно упрощены по сравнению с распознаванием речи полностью.

Подзадачей распознавания является также отделение<sup>10</sup> записи речи от прочих звуков (шумы, наложения) в записи, выделение речи на общем фоне, разделение источников речи, если их звучит несколько одновременно<sup>11</sup>.

Наконец, решается также подзадача автоматического определения качества речи<sup>12</sup>, востребована в системах связи, хранения речи, для задач синтеза речи.

**Б) Распознавание эмоций в речи**<sup>13</sup>. Важная информация, которая передается голосом, — это не только содержание речи, но и настроение говорящего. Задача распознавания настроения востребована в бизнесе (определение удовлетворения от оказанных услуг), здравоохранении, безопасности (определение наличия угрозы исходящей от говорящего) и др.

**В) Распознавание диктора**<sup>14</sup>. Помимо распознавания содержания речи востребованы задачи распознавания принадлежности речи:

- идентификация диктора<sup>15</sup> — определение того, на голос какой персоны более всего похож проверяемый голос;
- верификация диктора<sup>16</sup> — проверка того, что проверяемый голос принадлежит заданной персоне (например, при авторизации в онлайн-сервисах по голосу);
- разделение<sup>17</sup> и диаризация<sup>18</sup> дикторов — определение того, какой из дикторов говорит сейчас, выделение частей записи, принадлежащих одному диктору (в том

<sup>8</sup> См. <https://paperswithcode.com/task/accented-speech-recognition>

<sup>9</sup> См. <https://paperswithcode.com/task/keyword-spotting>

<sup>10</sup> См. <https://paperswithcode.com/task/speech-extraction>

<sup>11</sup> См. <https://paperswithcode.com/task/speech-separation>

<sup>12</sup> См. <https://paperswithcode.com/task/speech-quality>

<sup>13</sup> См. <https://paperswithcode.com/task/speech-emotion-recognition>

<sup>14</sup> См. <https://paperswithcode.com/task/speaker-recognition>

<sup>15</sup> См. <https://paperswithcode.com/task/speaker-identification>

<sup>16</sup> См. <https://paperswithcode.com/task/speaker-verification>

<sup>17</sup> См. <https://paperswithcode.com/task/speaker-separation>

<sup>18</sup> См. <https://paperswithcode.com/task/speaker-diarization>



числе без идентификации самого диктора), также в случаях, когда дикторы говорят одновременно;

- оценка физического состояния дикторов<sup>19</sup> — голос несет в себе некоторую информацию о физическом и эмоциональном состоянии диктора, его речевого аппарата. Оценка такого состояния может быть полезна для наблюдения за состоянием здоровья, например на сложных производствах, выдаче допуска к работе.

Г) **Моделирование и представление речи.** Востребованы задачи создания акустических моделей речи<sup>20</sup>, способов ее представления, выделение базовых элементов речи<sup>21</sup>, которые помогали бы лучше решать задачи распознавания и синтеза. Следует отметить задачу создания векторных представлений речи с помощью методов самообучения<sup>22</sup>, для которых не требуются заранее распознанные наборы данных.

### 3. ЗАДАЧИ СИНТЕЗА РЕЧИ

Задачи синтеза речи<sup>23</sup> следует различать по виду источника информации о синтезируемой речи, ее содержании.

А) **Синтез речи из текста**<sup>24</sup> (text-to-speech). Это наиболее популярная постановка задачи синтеза, когда есть текст, который следует озвучить голосом (или голосами). Дополнительно могут использоваться специальные значки, регулирующие эмоциональную окраску<sup>25</sup> речи, выразительность речи<sup>26</sup> и другие ее характеристики (произношение, акцент и т. п.). Также может быть востребована задача исправления произнесенной и записанной речи, удаление части произнесенного текста (относительно простая задача) или добавление в запись речи необходимого текста (более сложная задача)<sup>27</sup>.

Синтез речи широко используется в различных областях, например в голосовых помощниках, роботах-ассистентах, голосовых чат-ботах, искусстве.

Б) **Синтез речи из речи** (speech-to-speech). Здесь источником синтезированной речи является другой речевой сигнал, который преобразуется заданным образом. Например, улучшают качество звучания речи<sup>28</sup>,

<sup>19</sup> См. <https://paperswithcode.com/task/speaker-profiling>

<sup>20</sup> См. <https://paperswithcode.com/task/acoustic-modelling>

<sup>21</sup> См. <https://paperswithcode.com/task/unsupervised-mnist>

<sup>22</sup> См. <https://paperswithcode.com/task/acoustic-unit-discovery>

<sup>23</sup> См. <https://paperswithcode.com/task/speech-synthesis>

<sup>24</sup> См. <https://paperswithcode.com/task/text-to-speech-synthesis>

<sup>25</sup> См. <https://paperswithcode.com/task/emotional-speech-synthesis>

<sup>26</sup> См. <https://paperswithcode.com/task/expressive-speech-synthesis>

<sup>27</sup> См. <https://paperswithcode.com/task/speech-editing>

<sup>28</sup> См. <https://paperswithcode.com/task/speech-enhancement>

убирают шумы<sup>29</sup>, эхо<sup>30</sup>, помехи<sup>31</sup> и искажения, изменяют характеристики голоса (мужской в женский и наоборот, переделка тембра и т. п.), удаляют звуки-паразиты и др.

- В) **Синтез речи из изображения** (image-to-speech). Одна из задач — это озвучить содержание некоторого изображения для слабовидящих людей. Может решаться как через перевод изображения в текст его аннотации, а затем озвучивание текста, так и напрямую. Аналогичная задача перевода видео в речь, например чтение и озвучивание движений губ диктора.
- Г) **Синтез речи из мультимодальных источников**, когда используют несколько разных источников для создания речи. Например, задача клонирования голоса, где требуется воспроизвести заданный текст-источник заданным же голосом-источником, который изначально такой текст не озвучивал, или переделать голос, озвучивающий некоторый текст, на заданный голос<sup>32</sup>.
- Д) **Синтез речи по другим источникам**. Речь может создаваться и из других источников, например из электромиограммы. В некоторых задачах, например для создания произведений искусства или развлечения, достаточно создать звуки, похожие на голос человека, но не имеющие смысла, симитировать речь человека. Тогда источником псевдо-речи являются случайные числа или некоторые действия пользователя (скажем, движения мыши, нажатие клавиш). Сюда же отнесем озвучивание инопланетных персонажей игр или фильмов и вообще синтез нечеловеческой речи.

#### 4. АРХИТЕКТУРЫ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ПРЕДСТАВЛЕНИЯ И РАСПОЗНАВАНИЯ РЕЧИ

Речь представляется последовательностью во времени некоторых векторов (амплитуд звукового сигнала или коэффициентов спектра), значит, для обработки ее необходимо применять методы, работающие с последовательностями.

Два больших класса нейронных сетей<sup>33</sup> используются для обработки последовательностей — это *рекуррентные сети* и *сети на основе механизма внимания*, которые часто используются вместе.

##### 4.1. Рекуррентная нейронная сеть

В рекуррентной сети нейроны или слои нейронов в качестве входов принимают также и предыдущие, запомненные с прошлых тактов по времени, выходы нейронной сети или других ее слоев. В рекуррентных сетях, полученных из обычных полносвязных слоев нейронов, затруднен расчет градиента функции ошибки по параметрам, который необходим для обучения сети. Градиент сильно затухает (или,

<sup>29</sup> См. <https://paperswithcode.com/task/speech-denoising>

<sup>30</sup> См. <https://paperswithcode.com/task/acoustic-echo-cancellation>

<sup>31</sup> См. <https://paperswithcode.com/task/speech-dereverberation>

<sup>32</sup> См. <https://paperswithcode.com/task/voice-conversion>

<sup>33</sup> Базовое описание слоев нейронных сетей и их работы см. в разделе XXX.5.2





наоборот, сильно возрастает) с течением времени, при переходе к последующим тактам. Для борьбы с этим явлением используют *вентильные рекуррентные нейронные сети*: LSTM, GRU и подобные.

В вентильных сетях коэффициенты забывания информации с предыдущих тактов времени, равно как и коэффициенты добавления новой информации для текущего такта, регулируются за счет использования вентилей — умножителей. Вентиль забывания получает входную информацию и умножает ее на коэффициент от 0 до 1, тем самым имитируя забывание. Этот коэффициент сам по себе является обучаемым — это выход обычного полносвязного нейрона с функцией активации типа сигмоида. Изменяя в процессе обучения весовые коэффициенты такого нейрона, изменяется (т. е. регулируется) и коэффициент забывания. Если коэффициент получается близким к единице — прошлая информация практически не забывается, градиент будет относительно большим. Аналогично устроены вентили для добавления новой информации с текущего такта<sup>34</sup>. Различными комбинациями вентилей получают различные вентильные сети, так в классических LSTM сетях используется четыре разных обучаемых нейрона для вентилей, в GRU 3. Существуют тысячи модификаций таких вентильных сетей [1]. Отметим двунаправленные рекуррентные сети, в которых объединяется информация обработки последовательности слева-направо и справа-налево, это позволяет улучшить обработку в случаях, когда есть связь текущих отсчетов с последующими (а для текста или звука речи это, очевидно, так), например Bi-LSTM<sup>35</sup> сети.

#### 4.2. Механизм внимания

В рекуррентных сетях влияние отсчетов последовательности друг на друга последовательное, первый влияет на второй, второй на третий и т. д. Но на практике это далеко не всегда так. Некоторые отсчеты не влияют на другие вообще (например, тишина в звуке), некоторые оказывают сильное влияние только на соседние отсчеты, а другие — на далеко отстоящие отсчеты. Хотелось бы моделировать такое влияние. Для этого предложен механизм внимания [2]: для каждого отсчета рассчитывают, насколько сильно он связан со всеми другими отсчетами. Этот механизм обучаемый, поэтому можно регулировать за счет обучения влияние отсчетов друг на друга. Такое влияние отсчетов (внимание, attention) можно устанавливать как между различными последовательностями, так и для отсчетов одной последовательности<sup>36</sup> (самовнимание, self-attention). Вариант реализации механизма самовнимания вкратце: элементы входной последовательности (это векторы) путем умножения на обучаемые матрицы преобразуются в три вектора: Ключ,

<sup>34</sup> Детальное описание и визуализация архитектуры LSTM и GRU сетей: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>  
<https://towardsdatascience.com/illustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf21>

<sup>35</sup> Визуализация bi-LSTM сетей <https://analyticsindiamag.com/complete-guide-to-bidirectional-lstm-with-python-codes/>

<sup>36</sup> Визуализация работы механизма самовнимания: <https://towardsdatascience.com/illustrated-self-attention-2d627e33b20a#8481>

Запрос и Значение. Находят скалярные произведения Запроса одного отсчета со всеми Ключами других (включая текущий) отсчетов, с помощью softmax активации преобразуют эти числа в числа в диапазоне от 0 до 1 — это степени влияния отсчетов на текущий. Вектора Значений всех отсчетов умножаются на эти степени влияния и складываются, получается текущий элемент выходной последовательности. Так повторяют для всех отсчетов входной последовательности. Такой механизм преобразует входную последовательность в выходную, в которой величины влияния отсчетов учтены, причем матрицы преобразований обучаемые, могут подстраиваться под наборы обучающих данных. Затем уже эти преобразованные последовательности используют для решения задач распознавания или синтеза.

На основе механизма внимания создана архитектура нейронной сети Transformer<sup>37</sup>, состоящая из двух связанных частей: кодера и декодера. В процессе обучения кодер принимает входную последовательность (например, текст на одном языке), обрабатывает ее механизмом самовнимания, полученная последовательность поступает в декодер. Декодер принимает выходную последовательность (например, текст на другом языке, она известна при обучении), обрабатывает ее механизмом самовнимания, полученную последовательность обрабатывает механизмом взаимного внимания с выходом кодера. Для результирующей последовательности вычисляются уровни уверенности токенов (значений отсчетов), которые при обучении должны иметь максимальное значение для тех токенов, которые действительно встретились в выходной последовательности на текущем отсчете. После обучения такая архитектура может предсказывать наиболее вероятный токен выходной последовательности и тем самым строить ее. Показала эффективную работу в задачах обработки текста, звуков и др.

### 4.3. Векторные представления речи. Архитектура Wav2Vec

Работа с исходными представлениями речи в виде звуковой волны или спектрограммы напрямую может быть затруднена, требуются довольно сложные архитектуры для распознавания и синтеза речи. Возможно создавать новые, векторные представления речи, которые лучше подходят для ее обработки. Один из вариантов — создавать такие представления в процессе самообучения нейронной сети. Такой подход реализован в архитектуре Wav2Vec<sup>38</sup> [3].

Речевой сигнал в форме звуковой волны сверточным слоем преобразуется в некоторую последовательность  $Z$ . Эта последовательность поступает в кодер — часть сети Transformer, где преобразуется в выходную последовательность  $C$ . Одновременно последовательность  $Z$  квантуется по известным правилам в последовательность  $Q$ . В процессе самообучения часть элементов последовательности  $Z$  скрывается (маскируется), и в процессе самообучения ставится задача по оставшимся элементам  $Z$  получить последовательность  $C$ , как можно лучше совпадающую с  $Q$ .

Важный механизм здесь — это квантование  $Z$  в  $Q$ , он использует набор кодовых книг, каждая из которых содержит несколько обучаемых кодовых векторов. Для каждого вектора из текущего отсчета  $Z$  среди кодовых векторов каждой кодовой книги ищется наиболее похожий, отображенные векторы из каждой кодовой книги

37 См. <https://medium.com/inside-machine-learning/what-is-a-transformer-d07dd1fbec04>

38 См. <https://neurosys.com/wav2vec-2-0-framework/>



контактенируются и линейным слоем проецируются в один вектор, который и будет соответствующим элементом  $Q$ . Кодовые векторы, как и линейный слой, — обучаемые, подбираются в процессе обучения. Чтобы кодовые векторы были разнообразны, не вырождались, в функцию ошибки при обучении добавляют член, максимизирующий такое разнообразие (читай, энтропию).

Таким образом, создается новое векторное представление  $S$  для звука речи, которое затем уже используется для решения задач обработки, например распознавания.

## 5. НЕЙРОННЫЕ СЕТИ ДЛЯ СИНТЕЗА РЕЧИ ИЗ ТЕКСТА

При синтезе звука речи из текста исходными данными является текст — последовательность некоторых символов. Это могут быть символы алфавита, специальные фонетические значки или даже некоторые управляющие символы, описывающие, например, эмоциональную окраску произносимой речи. Часто такие символы переводят за счет обучения в векторное представление, тогда входом для системы синтеза будет последовательность таких векторов. Существует около десятка наборов данных с примерами записанной речи и транскрипцией<sup>39</sup> ее, на разных языках, в разных условиях. Результат синтеза — это звуковая волна, описывающая синтезированную речь. На сегодня по качеству синтезируемой речи лидируют два подхода, архитектура Tacotron и WaveNet<sup>40</sup>, но исследуются и другие подходы, связанные с использованием генеративно-сопоставительных сетей или скрытых марковских моделей [4]. Оценка качества синтезируемой речи производится опросом людей о естественности звучания, оцениваемом обычно по пятибалльной шкале. Отметим, что в таких опросах оценка естественной речи более 4.5 балла (для английского языка).

### 5.1. Архитектура Tacotron

Архитектура Tacotron [5] для синтеза речи из текста состоит из кодера, декодера на основе механизма внимания и сети постобработки, принимает символы текста и формирует кадры спектрограммы, которые затем трансформируются в звуковую волну.

В кодере входная последовательность (это последовательность унитарных векторов символов текста) проходит несколько полносвязных слоев. Результат — некоторое векторное представление  $V$  элементов последовательности. Последовательность  $V$  сворачивается набором (банком) одномерных сверток со сдвигом по оси времени 1 (не изменяется временное разрешение) и изменяемой (по оси времени) шириной от 1 до  $K$  (показывает количество учитываемых соседних отсчетов). Результаты этого объединяются и дополнительно слоем субдискретизации по оси

<sup>39</sup> См. <https://paperswithcode.com/datasets?task=speech-synthesis>

<sup>40</sup> См. <https://paperswithcode.com/sota/speech-synthesis-on-north-american-english>

времени уменьшают количество отсчетов. Полученная последовательность проходит несколько сверточных слоев с заданной шириной. К полученной последовательности добавляют последовательность  $V$ , пропускают результат через несколько полносвязных слоев и, наконец, слой двунаправленной рекуррентной сети GRU. Таким образом, кодер формирует другое векторное представление текста.

Выходная последовательность (здесь это полосы мел-спектра звука речи, соответствующей произносимому тексту) проходит через декодер. Процесс обучения концептуально похож на работу Transformers. Элементы выходной последовательности сначала проходят через полносвязные слои, формируя скрытое векторное представление. Эти векторы поступают в рекуррентный GRU слой, который формирует последовательность запросов  $Q$ , которая проходит через слой взаимного внимания с выходом кодера. Результат слоя внимания вместе с  $Q$  поступает на следующие GRU слои. Полученная из них последовательность проходит через блок слоев, похожий на кодер (банк сверток, субдискретизация и др.).

Выход всей этой сети интерпретируется как последовательность коэффициентов полос мел-спектра. Сеть обучается предсказывать следующую полосу (т. е. саму выходную последовательность, задержанную на один такт времени). При обучении выходная последовательность известна, поэтому сеть может обучаться предсказывать следующий ее элемент по предыдущему. После обучения выходная последовательность предсказывается пошагово, начинают с пустой выходной последовательности, в которую добавлен только элемент начала последовательности, на основе его предсказывают следующий, вставляют его в выходную последовательность и на основе уже двух элементов предсказывают третий и так далее. В результате получают мел-спектрограмму для синтезированного голоса, которая затем преобразуется в звуковую волну.

Для оценки качества синтеза был проведен опрос [5] носителей языка с целью оценить по пятибалльной шкале естественность синтезированного голоса (100 фраз), получили в среднем 3.82, для сравнения синтезатор на основе скрытых марковских моделей дал оценку в 4.09 балла.

## 5.2. Архитектура WaveNet

Tacotron переводит текст в спектрограмму, а уже спектрограмма переводится в звук речи. В этом месте теряется довольно заметная часть естественности звучания, появляются слышимые искажения. В архитектуре WaveNet [6] избавились от промежуточного шага в виде спектрограммы и переводят текст сразу в звук<sup>41</sup>.

Архитектура является управляемым генератором и состоит из двух частей. Первая часть — преобразователь, который изменяет последовательность звуковых фрагментов (отсчетов или групп отсчетов), связывает текущий с предыдущими. Это полностью сверточная нейронная сеть, которая принимает звуковую волну — последовательность фрагментов звука и возвращает преобразованную последовательность, в которой учтены связи с прошлыми фрагментами. Поскольку здесь не должно быть зависимости от будущих фрагментов, то используются каузальные свертки, когда для предсказания текущего используют только предыдущие

41 Анимацию и пояснение работы см.: <https://deepmind.com/blog/article/wavenet-generative-model-raw-audio>

фрагменты: для предсказания второго фрагмента используется только первый, для предсказания третьего фрагмента используется первый и второй и т. д. Для увеличения размера окна влияния фрагментов используют распределенные (dilated) свертки, когда в операцию свертки берут не все входы подряд, а через один, через два и т. п., т. е. пропускают некоторые входы. Сеть многослойная, используя распределенные свертки с постепенно возрастающим количеством пропускаемых отсчетов (например, но не обязательно, на первом слое брать каждый отсчет, на втором слое брать каждый второй, на третьем — каждый четвертый и т. д.), можно получить достаточно большой размер окна влияния предыдущих фрагментов на текущий.

Задача всей сети — предсказать следующий фрагмент. Для этого нужно ввести представление амплитуды и добавить управляющую информацию, например текст, от которого будет зависеть генерируемый звук. Обычный способ представления амплитуды потребовал бы предсказания уровня уверенности для каждого из возможных 65 536 значений (16-битный звук), поэтому здесь квантуют значения амплитуды по нелинейному закону до 256 разных состояний, что существенно уменьшает размеры сети.

Вторая часть сети, блок управления, в котором добавляется управляющая информация.

Этот блок основан на использовании последовательного набора обучаемых вентилях (как в вентильных рекуррентных сетях, но модифицированных для сверточных слоев). Набор состоит из нескольких вентилях, каждый из которых принимает выход  $x$  предыдущего (первый вентиль принимает выход преобразователя), пропускает его через два обучаемых сверточных слоя с активациями гипертангенс (изменяется от -1 до 1) и сигмоида (изменяется от 0 до 1), результаты которых умножаются в вентиле. Выход вентиля проходит через свертку с ядром 1 на 1 (т. е. сворачиваются только каналы), результат которой поступает на основной выход набора. Также этот выход складывается с входом  $x$  (как в сетях ResNet) и поступает на вход следующего вентиля. Основные выходы всех вентилях проходят через ряд сверточных слоев и в итоге через слой с softmax активацией, который и предсказывает уровень уверенности для квантованных амплитуд звукового фрагмента.

Для управления генератором в блок вентилях добавляют еще один (обучаемый) вход  $h$ . Он может быть как локальным — последовательностью чисел такой же частоты дискретизации, что и звуковой сигнал, так и менее локальным, с меньшей частотой. В последнем случае применяют транспонированные свертки, чтобы получить управляющую последовательность той же частоты дискретизации, что и звук. Сама управляющая последовательность может представлять собой различную информацию: текст для его озвучивания; характеристики голоса для его преобразования (клонирования) и др.

В результате получилась довольно точная и быстрая сеть для синтеза речи, в работе [6] опрос по пятибалльной шкале о естественности

синтезируемой речи дал 4.21 балла для североамериканского и 4.08 для китайского мандаринского наречий. Для сравнения, запись естественной речи по тому же опросу показала 4.55 и 4.21 соответственно.

Сочетание идей сети Tacotron, которая синтезирует мел-спектрограмму по тексту, и сети WaveNet, которая использует эту спектрограмму как управляющую последовательность для синтеза звука, привело к появлению сети Tacotron 2 [7], которая показала качество синтеза в 4.526 балла (опрос по пятибалльной шкале).

## 6. НЕЙРОННЫЕ СЕТИ ДЛЯ СИНТЕЗА ПЕВЧЕСКОГО ГОЛОСА

Поющие автоматы привлекают внимание исследователей в области синтеза голоса, но для синтеза певческого голоса предъявляются повышенные требования к качеству синтеза, соответствия голоса песне, контролю высоты и продолжительности звучания и т. п.

### 6.1. Модели на основе процесса диффузии

В [8] предложено использовать модель диффузии для создания певческого голоса, которая показала хорошие результаты<sup>42</sup>. Модель диффузии предполагает, что мы рассматриваем процесс синтеза как случайный. Исходное распределение входных векторов можно постепенно, небольшими шагами, зашумлять, пока не получится шумоподобное распределение. Здесь используется Гауссовский процесс зашумления<sup>43</sup>. При разумных ограничениях на процесс зашумления возможен обратный процесс — из зашумленных образцов получать очень похожие на исходные. При реализации обратного процесса нам не известна функция для перевода распределений, поэтому ее моделируют некоторой обучаемой моделью, нейронной сетью, подбирая параметры в процессе обучения. Тогда такая модель может выступать как генератор образцов, похожих на настоящие, для нашего примера подстроится под конкретный певческий голос. Добавив в модель управляющий сигнал (что именно петь), получим синтезатор певческого голоса.

Эта модель DiffSinger реализована в [8], она оперирует с мел-спектрограммами в модели диффузии. В процессе работы такой системы сначала получают мел-спектрограмму желаемого голоса каким-либо известным способом (в работе это декодер на основе Transformer архитектуры [9]) из текста и музыкальных характеристик голоса (высота, длительность звучания). Такую мел-спектрограмму можно считать сильно зашумленным образцом, который не обладает достаточным качеством звучания. Затем, начиная с этой мел-спектрограммы, запускают обратный процесс диффузии (который был, разумеется, уже обучен), получая итеративно все более и более качественные мел-спектрограммы для голоса. В качестве нейронной сети для убирания шума на этих шагах используют вариант сети WaveNet (не каузальная и работающая с мел-спектрограммами). Наконец, полученную мел-спектрограмму переводят в звуковую волну (здесь используют предобученную сеть Parallel WaveGAN [10]). В результате получили (опрос по пятибалльной шкале) качество синтеза 3.85 балла. Без процесса диффузии качество было 3.71 балла. Ближайший

<sup>42</sup> Прослушать можно здесь: <https://diffsinger.github.io/>

<sup>43</sup> Визуализацию см. <https://lilianweng.github.io/lil-log/2021/07/11/diffusion-models.html>



аналог, архитектура на основе генеративно-состязательной сети, показала качество 3.74 балла, тогда как естественная запись голоса оценена в 4.30 балла и 4.04, если эта запись восстанавливалась из её мел-спектрограммы. Последнее сравнение показывает, что мел-спектрограммы действительно дают заметную потерю качества звучания. Справедливости ради отметим, что для обычных людей на слух достигнутое качество синтеза вполне приемлемо.

## 6.2. Модели на основе генеративно-состязательных сетей

Синтез речи — это задача генерации. Для генерации объектов из заданного распределения и собственно получения такого распределения хорошо зарекомендовали себя генеративно-состязательные сети. В таких сетях есть две соревнующиеся стороны: Генератор и Дискриминатор. Генератор учится создавать примеры образцов. Дискриминатор же учится различать настоящие образцы и сгенерированные Генератором. Цель Генератора — обмануть Дискриминатор, создать такой образец, что Дискриминатор не сможет его отличить от настоящего. Постепенно и попеременно обучая Генератор и Дискриминатор, можно добиться генерации правдоподобных образцов. Такой подход может быть применен для создания образцов различной природы, в том числе и голоса.

В работе [11] предложена генеративно-состязательная сеть Multi-Singer для реализации певческого голоса, которая обучалась на наборе данных с несколькими певцами.

Генератор сделан на основе модификаций блоков из сети WaveNet, принимает мел-спектрограмму синтезируемого голоса и вектор шума, генерирует звуковые волны, представленные в разных частотных полосах (две высокой и две низкой), и объединяет их в одну звуковую волну.

Дискриминатор модели основан на LSTM сети, принимает звуковую волну (от Генератора или от настоящей записи) и вектор характеристик голоса певца. Возвращает, очевидно, уровень уверенности в том, является ли проверяемая звуковая волна настоящей, принадлежащей заданному певцу, или сгенерированной. В работе используют второй Дискриминатор, который не учитывает характеристики голоса певца, а только лишь пытается отличить настоящую звуковую волну (не важно, кому именно принадлежащую) от сгенерированной. Этот дискриминатор состоит из сверточных слоев.

Функция ошибки общая, учитывает ошибки первого и второго дискриминаторов. Для обучения Генератора используют дополнительные члены в функцию ошибки, которые: а) сравнивают характеристики голоса певца по мел-спектрограмме для сгенерированной и настоящей звуковой волны и б) коэффициенты оконного преобразования Фурье для настоящей и сгенерированной звуковой волны.

Мел-спектрограммы для Генератора берутся из модификации сети FastSpeech 2 [9] на основе Transformers, которая создает мел-спектрограммы из

текста и музыкальных характеристик голоса. Характеристики голоса певца берутся из предобученной сети-векторизатора на основе LSTM сети [12].

В результате получилась довольно хорошая архитектура для создания певческого голоса<sup>44</sup>, опрос по пятибалльной шкале о естественности звучания показал 3.96 для известных и 3.98 для неизвестных певцов, естественные записи оценивались в 4.03 балла.

## 7. ЗАКЛЮЧЕНИЕ

Сегодня исследователи применений нейронных сетей для синтеза голоса рассматривают совершенно различные подходы, часто комбинируемые вместе. Это и архитектуры на основе рекуррентных сетей, и на основе механизма внимания, и генеративные сети и подходы на основе процесса диффузии и другие. Используют как представление в виде звуковой волны, так и мел-спектрограммы. Качество синтезируемого голоса практически достигло уровня человеческого восприятия. Появились синтезаторы певческих голосов, также с довольно высоким качеством звучания. Однако следует отметить, что такие высокие результаты получены для небольшого набора языков, английского и китайского, для других языков результаты гораздо скромнее. В основном это связано с недостаточным качеством наборов обучающих данных и вычислительными ресурсами, которые необходимо затратить на обучение моделей. Например, для русского языка на наборе данных RUSLAN была получена естественность звучания 3.78 (по пятибалльной шкале) в задаче синтеза голоса по тексту [13].

## Литература

1. Jozefowicz, R., Zaremba, W., & Sutskever, I. (2015, June). An empirical exploration of recurrent network architectures. In International conference on machine learning (pp. 2342–2350). PMLR.
2. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In Advances in neural information processing systems (pp. 5998–6008).
3. Baevski, A., Zhou, H., Mohamed, A., & Auli, M. (2020). wav2vec 2.0: A framework for self-supervised learning of speech representations. arXiv preprint arXiv:2006.11477.
4. Tan, X., Qin, T., Soong, F., & Liu, T. Y. (2021). A survey on neural speech synthesis. arXiv preprint arXiv:2106.15561.
5. Wang, Y., Skerry-Ryan, R. J., Stanton, D., Wu, Y., Weiss, R. J., Jaitly, N., ... & Saurous, R. A. (2017). Tacotron: Towards end-to-end speech synthesis. arXiv preprint arXiv:1703.10135.
6. Oord, A. V. D., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., ... & Kavukcuoglu, K. (2016). Wavenet: A generative model for raw audio. arXiv preprint arXiv:1609.03499.
7. Shen, J., Pang, R., Weiss, R. J., Schuster, M., Jaitly, N., Yang, Z., ... & Wu, Y. (2018, April). Natural tts synthesis by conditioning wavenet on mel spectrogram predictions. In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 4779–4783). IEEE.
8. Liu, J., Li, C., Ren, Y., Chen, F., Liu, P., & Zhao, Z. (2021). DiffSinger: Singing Voice Synthesis via Shallow Diffusion Mechanism. arXiv preprint arXiv:2105.02446.
9. Ren, Y., Hu, C., Tan, X., Qin, T., Zhao, S., Zhao, Z., & Liu, T. Y. (2020). FastSpeech 2: Fast and high-quality end-to-end text to speech. arXiv preprint arXiv:2006.04558.
10. Ryuichi Yamamoto, Eunwoo Song, and Jae-Min Kim. Parallel wavegan: A fast waveform generation model based on generative adversarial networks with multi-resolution spectrogram. In ICASSP

<sup>44</sup> Прослушать: <https://multi-singer.github.io/>





- 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 6199–6203. IEEE, 2020
11. Huang, R., Chen, F., Ren, Y., Liu, J., Cui, C., & Zhao, Z. (2021, October). Multi-Singer: Fast Multi-Singer Singing Voice Vocoder With A Large-Scale Corpus. In Proceedings of the 29th ACM International Conference on Multimedia (pp. 3945–3954).
  12. Wan, L., Wang, Q., Papir, A., & Moreno, I. L. (2018, April). Generalized end-to-end loss for speaker verification. In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 4879–4883). IEEE.
  13. Gabdrakhmanov, L., Garaev, R., & Razinkov, E. (2019, August). Ruslan: Russian spoken language corpus for speech synthesis. In International Conference on Speech and Computer (pp. 113–121). Springer, Cham.

## NEURAL NETWORKS FOR SPEECH SYNTHESIS OF VOICE ASSISTANTS AND SINGING MACHINES

*Pantiukhin D.V., Senior Lecturer, National Research University Higher School of Economics, Moscow, Russia  
Senior Lecturer, MIREA – Russian Technological University, Moscow, Russia, dpantiukhin@hse.ru*

The paper presents an overview of recent publications on the use of neural network methods for speech synthesis. A brief introduction to neural network architectures for speech synthesis is given. Methods for creating a singing voice for voice assistants and singing machines are considered.

• *neural networks* • *singing machines* • *voice synthesis*.

# Определение частоты основного тона речевого сигнала на основе индекса многомерной синхронизации

*Харченко С.С., кандидат технических наук, доцент кафедры безопасности информационных систем Томского государственного университета систем управления и радиоэлектроники, г. Томск, Россия, kss@fb.tusur.ru*

*Крючков И.А., разработчик программного обеспечения отдела разработки компонентов АО «Инфотекс», г. Томск, Россия*

*Мещеряков Р.В., доктор технических наук, профессор РАН, главный научный сотрудник лаборатории киберфизических систем ИПУ РАН, Москва, Россия, mr@ieee.org*

Статья посвящена исследованию определения частоты основного тона речевого сигнала. Материалом для исследования послужила база данных The Pitch-Tracking Database from Graz University of Technology [33]. В работе использовался разработанный алгоритм определения частоты основного тона речевого сигнала на основе индекса многомерной синхронизации. Для интерпретации полученных результатов использовалось сравнение с известными эстиматорами, в качестве метрик использовались процент грубых ошибок и средний процент мелких ошибок. В ходе исследования было установлено, что алгоритм целесообразнее использовать для определения средней частоты основного тона, чем для определения частоты основного тона в режиме реального времени.

• частота основного тона • индекс многомерной синхронизации • речевой сигнал.

## ВВЕДЕНИЕ

Конец XX и начало XXI в. ознаменовались началом истории развития речевых технологий. За это время было решено множество фундаментальных прикладных задач и разработано большое количество разнообразных алгоритмов обработки речевых сигналов, что, впрочем, не вызвало падения интереса к этой теме [1]. Активные работы как по улучшению существующих, так и по разработке новых алгоритмов обработки речи ведутся по сей день.



Одним из основных параметров, на использовании которого базируется как большинство классических, так и недавно разработанных алгоритмов оценки, распознавания и синтеза речевых сигналов, является частота основного тона (ЧОТ) [2].

Частота основного тона — это частота повторения колебаний голосовых складок при произнесении вокализованных звуков речи. Среди оценок данной частоты принято выделять мгновенное значение частоты основного тона и среднее значение частоты основного тона [2].

Процесс выделения ЧОТ как правило состоит из нескольких этапов [1]:

- 1) фильтрация и подавление шума,
- 2) сегментация на информативные участки,
- 3) определение информативных параметров.

Для каждого этапа может применяться отдельный алгоритм обработки речевого сигнала, что ведет к появлению гибридных модификаций на основе более простых алгоритмов. На данный момент наиболее популярными алгоритмами оценки ЧОТ являются сингулярный анализ [3–4], RAPT, YIN, SWIPE и их модификации [5], обеспечивающие низкий процент ошибок даже при наличии шумов. Однако даже столь совершенные алгоритмы не обеспечивают безошибочную работу во всех областях деятельности, связанных с обработкой сигналов. Например, при модуляции ЧОТ [5] точность определения заметно снижается, что и побуждает искать новые алгоритмы и улучшать существующие.

На практике оценка ЧОТ находит свое применение в самых различных областях человеческой деятельности. Помимо очевидного применения в идентификационных устройствах распознавания голоса операции детектирования и манипуляций с ЧОТ могут использоваться и в отрыве от речевого сигнала. Например, алгоритмы определения ЧОТ находят свое применение в музыке [6] или для определения переработанных с целью обхода авторских прав аудиоконпозиций [7].

Особенно широкое применение характеристики ЧОТ находят в медицине. Так, в работе Д.Д. Полешникова [8] продемонстрирован пример того, как с помощью параллельного анализа средней частоты тона речевого сигнала пациента и динамики изменения ЧОТ можно определить частоту сердечных сокращений пациента. Аналогичное исследование метрик ЧОТ может быть также использовано для диагностики повреждения наружной ветви верхнего гортанного нерва в тиреоидной хирургии [9]. Также определение ЧОТ может быть использовано при отслеживании рефлекторных изменений в процессе голосообразования у пациентов после ларингоэктомии [10]. Существуют специально отобранные и модифицированные алгоритмы для решения еще более специфических медицинских задач. Например, адаптивный метод разложения сигнала на эмпирические моды [11], использующийся для предварительной обработки звуковых сигналов с целью повышения

точности определения ЧОТ, нашел свое применение в системах диагностики пограничных психических расстройств [12].

Таким образом, точное определение частоты основного тона является одним из основных направлений в сфере обработки речевых сигналов. Что касается тенденций самого направления, можно выделить несколько основных спектров работ:

- 1) развитие алгоритмов выделения ЧОТ в режиме реального времени [13–14];
- 2) разработка алгоритмов с повышенной устойчивостью к шумам [15–16], в том числе с использованием искусственных нейронных сетей (ИНС) [17];
- 3) непосредственно обучение ИНС [18–22].

Обучение ИНС не на непосредственный анализ речи, а на выделение частоты основного тона в ней дает более успешные результаты, например в случаях смешения голосов нескольких только мужских или только женских голосов в одном канале [23].

Обучение ИНС непосредственному анализу речи без выделения ЧОТ часто дает неудовлетворительные результаты. Например, при анализе мандаринского акцента ИНС без использования ЧОТ акцент распознавался лишь в 73 % случаев [24].

Стоит отметить, что базирующиеся на спектральном анализе методы обеспечивают большую точность, чем методы, которые базируются на использовании автокорреляционных преобразований [25]. Впрочем, существуют и гибридные алгоритмы, комбинирующие оба подхода [26]. В рамках данной работы проведено исследование эффективности выделения частоты основного тона с помощью метода многомерной синхронизации [29]. Сторонние исследования показывают, что даже двумерная обработка моноканального сигнала: представление сигнала в виде симметричной матрицы, позволяет повысить точность определения ЧОТ как в случае использования классических алгоритмов [30], так и при обучении ИНС [18].

В основе расчёта индекса многомерной синхронизации лежит автокорреляционный подход. Данный алгоритм отлично зарекомендовал себя в распознавании нейронной активности при использовании большого количества электродов для снятия сигнала, кроме того, демонстрирует высокую устойчивость к шумам [31].

Поскольку речь представляет из себя сложный полигармонический сигнал, предполагается, что индекс многомерной синхронизации может предложить увеличение точности выделения ЧОТ при обработке стерео канального речевого сигнала.

## МЕТОДЫ

За основу алгоритма был взят алгоритм выделения устоявшихся зрительно-вызванных потенциалов на основе индекса многомерной синхронизации [32]. Указанный алгоритм представляет собой модификацию классического корреляционного анализа.

1. Исходный сигнал дискредитируется и разбивается на участки определенной фиксированной длины.



2. Каждый из данных участков представляется в виде матрицы .
3. Выполняется построение списка матриц вида  $Yi = \begin{bmatrix} \sin(2\pi F_i t) \\ \cos(2\pi F_i t) \end{bmatrix}$  с гармониками той же длины (формула 1).
4. Для каждого участка исходного сигнала ( $Xi$ ) производится сравнение с каждой гармоникой из списка ( $Yi$ ) путем расчёта коэффициента корреляции данных матриц.
5. Частота  $F$ , для гармоник которой коэффициент корреляции дал наибольший результат, принимается за  $F0$  на данном участке  $Xi$ .

Различие состоит в том, что на четвертом этапе вместо коэффициента корреляции используется индекс многомерной синхронизации (выражение 2). Расчёт данного индекса представляет из себя последовательность операций с исходными матрицами, в результате которых они преобразуются в матрицу  $S$ , для которой производится расчёт вектора собственных значений, из которого и выводится итоговый коэффициент (выражение 6).

$$Y(t) = \dot{c} \tag{1}$$

где  $N$  — порядковый номер гармоники,  $K$  — количество отсчетов, —  $F_s$  частота дискретизации,  $Y(t)$  — опорный сигнал,  $X(t)$  — исходный сигнал.

$$S = \begin{bmatrix} I_{1 \times 1} & \frac{1}{\sqrt{D_{11}}} D_{12} \frac{1}{\sqrt{D_{22}}} \\ \frac{1}{\sqrt{D_{22}}} D_{21} \frac{1}{\sqrt{D_{11}}} & I_{2N_h \times 2N_h} \end{bmatrix} \tag{2}$$

где  $D_{xy}$  рассчитываются согласно выражениям 3–5

$$D_{11} = \left(\frac{1}{M}\right) X X^T \tag{3}$$

$$D_{22} = \left(\frac{1}{M}\right) Y Y^T \tag{4}$$

$$D_{21} = D_{12} = \left(\frac{1}{M}\right) X Y^T \tag{5}$$

$$R_i = 1 + \frac{\sum \lambda'_i \log(\dot{c} \lambda'_i)}{\log(\dot{c} P) \dot{c}} \dot{c} \tag{6}$$

где нормированные собственные значения  $S$ ,  $P = N + Nh$  и  $Nh$  обозначают число строк в опорном сигнале  $Y(t)$ . Частота фотостимуляции, определяется как частота, соответствующая максимальному значению  $R$ .

В случае анализа аудиосигнала исходные сэмплы  $Xi$  представляются в виде одномерных (монофонический сигнал) или двумерных (стереосигнал)

матриц. В качестве средства реализации исследуемого алгоритма был выбран высокоуровневый язык программирования Python 3.8. Благодаря простоте языка и большому количеству актуальных библиотек удалось значительно сократить время на программную реализацию алгоритма.

В качестве основного инструмента для проведения операций над матрицами была использована библиотека **numpy**. Для быстрого преобразования исходного аудиосигнала в массив отсчетов **numpy** была использована библиотека **librosa**. Для удобного вывода временных диаграмм использовался модуль **pyplot** библиотеки **matplotlib** в связке с модулем **display** библиотеки **librosa**. Для представления полученных результатов в виде таблицы Excel была использована библиотека **pandas**.

В качестве исходных данных для исследования была использована база данных **The Pitch-Tracking Database from Graz University of Technology** [33].

Эта база содержит 4718 сортированных вокализованных образцов речи 20 различных спикеров: 10 мужчин и 10 женщин. Особенность БД состоит в том, что в ней также содержатся измеренные с помощью ларингофона контрольные значения частоты основного тона  $F_0$ .

Исходный сигнал, полученный с помощью функционала библиотеки **librosa**, прежде чем подать его на вход алгоритма, предварительно нормализовался по амплитуде, а затем смещался вверх относительно временной оси. Таким образом на входе получался аудиосигнал той же формы, изменяющийся в пределах  $[0;2]$  и, таким образом, пригодный для использования алгоритмом, поскольку, как видно из выражения 2, в вычислении индекса используется операция взятия корня, что затрудняет анализ при наличии отрицательных значений в исходных данных.

Эксперимент проводился в следующих условиях.

1. Для расчёта  $F_0$  для дикторов разного пола применялись разные диапазоны исследуемых частот:
  - для дикторов мужского пола: 60–240 Гц;
  - для дикторов женского пола: 60–350 Гц.
2. Продолжительность одного нормализованного участка составляла  $t_{сэмпла} = 50$  мс.
3. Исходные данные были представлены в формате **“.wav”**.
4. Преобразование исходных данных в массивы отсчетов производилось на частоте дискретизации  $f_{дискретизации} = 48$ кГц.

В ходе предварительного изучения особенностей полученного алгоритма было обнаружено, что предложенный алгоритм обладает очень большой выборочностью относительно представленных данных: корреляция выявлялась только на участках, в которых стабильно (либо с незначительными колебаниями) присутствовала гармоника определенной частоты (рис. 1–2).

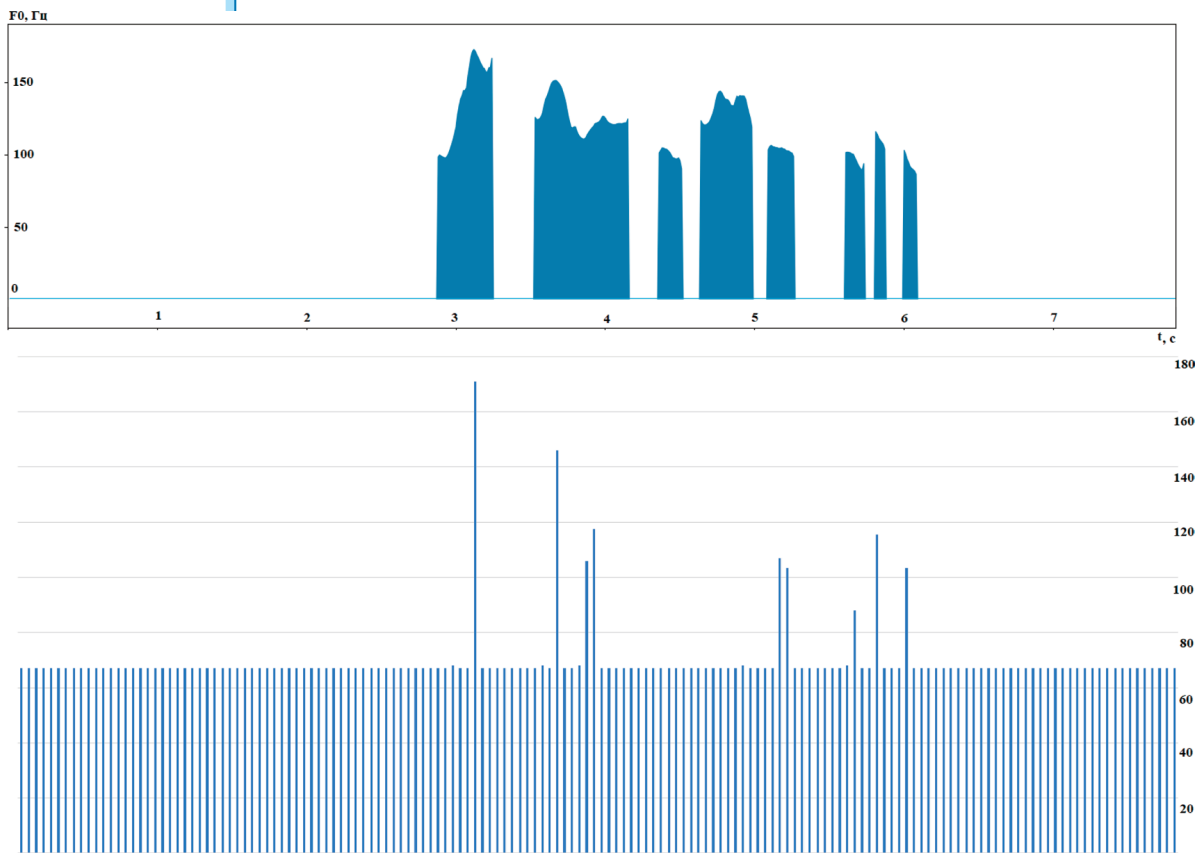


Рис. 1. Контрольные значения ларингофона и полученные с помощью разработанного алгоритма значения частоты основного тона

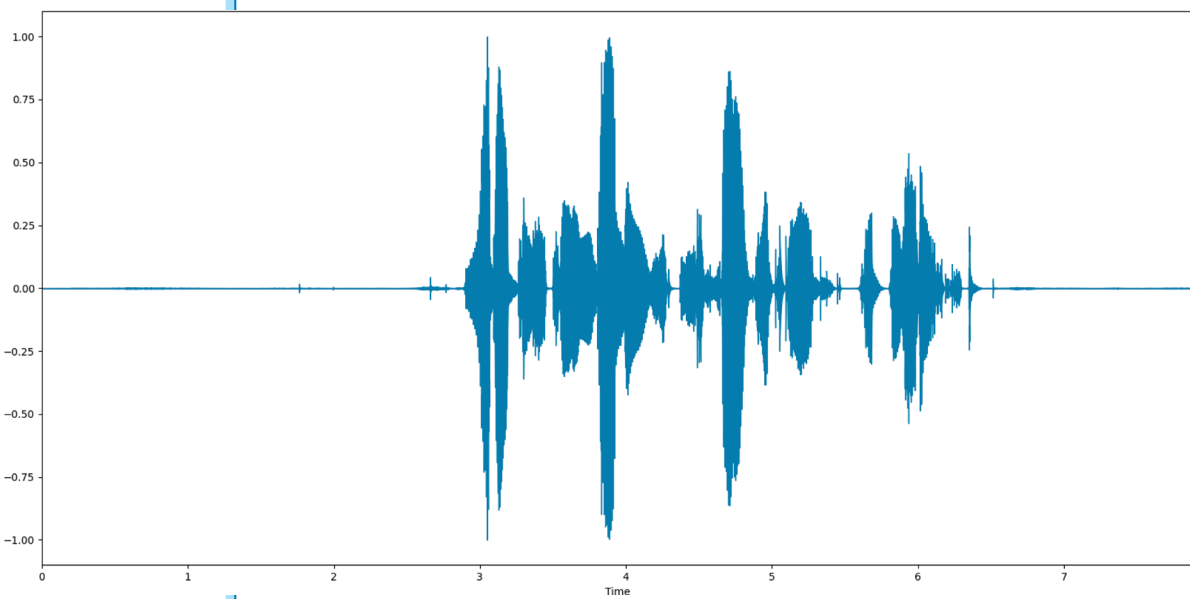


Рис. 2. Исходный сигнал

В связи с выделенной особенностью на эксперимент были наложены следующие дополнительные ограничения.

1. Расчёт статистических оценок производится только для фреймов, чья оценка превысила частотный порог в 70 Гц, т. е. только сэмплы, «распознанные» алгоритмом таким образом, считались вокализованными. Как правило, корреляция на «пустых» участках алгоритмом воспринималась одинаково и соответствовала определенной частоте в диапазоне 60–70 Гц. Данная «пустая» частота не зависела от предоставленных на вход образцов или уровня зашумленности, однако коррелировала с размером выбранного сэмпла: например при  $t_{\text{сэмпла}} = 50$  мс частота «отсутствия» корреляции находилась на уровне 67 Гц, что и можно наблюдать на рисунке 1.
2. В статистику принимались только те образцы, для которых алгоритм смог определить более одного участка корреляции выше указанного частотного порога.

Для сравнения алгоритмов выделения ЧОТ, как правило, рассматривается процент грубых ошибок GPE (gross pitch errors) [27] и средний процент мелких ошибок MFPE (mean fine pitch error) [28].

Величина GPE показывает отношение количества анализируемых фреймов с отклонением полученной оценки ЧОТ более чем на  $\pm 20\%$  от реального значения ЧОТ к общему числу вокализованных фреймов:

$$GPE(\%) = \frac{N_{GPE}}{N_V} 100,$$

где  $N_{GPE}$  — число фреймов с отклонением полученной оценки более чем на  $\pm 20\%$  от настоящего значения ЧОТ;

$N_V$  — общее число вокализованных фреймов.

20 %-я погрешность ошибки берется из соображений, что большинство ошибок, допускаемых алгоритмами при оценивании ЧОТ, варьируется в пределах октавы [27].

Величина MFPE показывает среднее значение отклонения оценки от действительного значения в процентах для всех оценок с отклонением менее  $\pm 20\%$ , т. е. для всех оценок, которые не являются грубыми ошибками:

$$MFPE(\%) = \frac{1}{N_{FPE}} \sum_{n=1}^{N_{FPE}} \frac{|F_0^{true}(n) - F_0^{est}(n)|}{F_0^{true}(n)} \times 100,$$

где  $N_{FPE}$  — число вокализованных фреймов без грубых ошибок;

$F_0^{true}(n)$  — действительные значения основного тона;

$F_0^{est}(n)$  — оценочные значения основного тона.



## РЕЗУЛЬТАТЫ И ДИСКУССИЯ

Данные, полученные в ходе проведения эксперимента, представлены в таблице 1.

Таблица 1

### Результаты использования MSI

Характеристика	Полученное значение	
	М	Ж
Полное время анализа базы данных	~ 6 часов	
Оборудование	AMD Ryzen 9 3900X 3.8GHz	
Среднее время анализа одного нормализованного участка (сэмпла)	15.2736 мс	21.129 мс
GPE	0.91245 %	1.697 %
MFPE	1.7305 %	2.113 %
Количество обработанных аудиозаписей	2360	2358
Количество аудиозаписей, принятых в статистику	2130	1660

Сравнительные данные алгоритма на основе индекса многомерной синхронизации (MSI) относительно других алгоритмов приведены в таблице 2.

Таблица 2

### Результаты использования алгоритмов [34]

	Мужской голос		Женский голос	
	GPE	MFPE	GPE	MFPE
RAPT	3.687	1.737	6.068	1.184
YIN	3.184	<b>1.389</b>	3.960	0.835
SWIPE'	<b>0.783</b>	1.507	4.273	<b>0.800</b>
IRAPT 1	1.625	1.608	3.777	0.977
IRAPT 2	1.571	1.565	3.777	1.054
MSI	0.912	1.731	<b>1,697</b>	2.113

Несмотря на то что исследование проводилось на моноакустических сигналах (в отсутствие базы стереосигнальной базы данных с подтвержденной ларингофоном частотой F0), алгоритм показывает весьма достойные результаты в сравнении с современными алгоритмами оценки ЧОТ, в частности — куда точнее определяет F0 на выборке дикторов женского пола, как видно из таблицы 2. Однако, как видно из таблицы 1, это обстоятельство обуславливается чрезвычайно низкой чувствительностью алгоритма: количество отброшенных данных для анализа достигает на той же выборке почти 30 %, чем, вероятно, и можно объяснить столь высокую точность.

Нельзя упускать из внимания и время обработки одного сэмпла: оно на порядки выше, чем у современных алгоритмов оценки ЧОТ, пригодных для использования в реальном времени [27], что может быть связано с средствами реализации, поскольку для каждого сэмпла необходимо рассчитать MSI относительно каждой гармоники на выбранном диапазоне частот. Если необходимо повысить точность, уменьшив шаг прохода по диапазону, количество вычислений возрастает пропорционально. По проведенным исследованиям можно сделать вывод, что алгоритм малоприменим для оценки траектории ЧОТ в реальном времени, однако может быть применен для аутентификации ЧОТ по уже существующему образцу: в этом случае недостаток алгоритма — необходимость проходить по каждой частоте для определения корреляции — становится одним из преимуществ, поскольку такая частота будет заранее известна в виде образца, и количество вычислений сокращается в сотни раз.

## ЗАКЛЮЧЕНИЕ

В связи с выявленной высокой точностью данного алгоритма, которую потенциально можно увеличить с использованием стереосигнала на входе, алгоритм может использоваться для уменьшения вероятности ошибки второго рода в системах биометрической идентификации и аутентификации.

Используя столь чрезвычайную выборочность алгоритма, его вполне можно применять и для оценки средней  $F_0$ , а также в ситуациях, где нет возможности отфильтровать исходный сигнал: как видно из таблицы 2, алгоритм показывает сравнимые результаты с другими современными алгоритмами оценки ЧОТ, несмотря на то что на входе им использовались сырые, необработанные от шумов данные.

В ходе подготовки эксперимента был обнаружен недостаток разработанного алгоритма при оценке данных дикторов мужского пола, связанный с кратными частотами: из-за выбранного диапазона частот в оценку алгоритма 15–20 %, под влиянием шумов, вносились небольшие погрешности, в результате в качестве  $F_0$  выбралась частота  $2 * F_{0 \text{ истинное}}$ , что в ходе эксперимента приходилось программно отлавливать: если частота была вдвое больше предыдущей, максимально коррелирующей с участком сигнала частоты и разница их оценок MSI не превышала установленный  $\epsilon = 0.001$ , то за  $F_0$  принималась первая частота.

## Литература

1. Алимуратов А.К. Обзор и классификация методов обработки речевых сигналов в системах распознавания речи / А.К. Алимуратов, П.П. Чураков — Измерение. Мониторинг. Управление. Контроль. — 2015. — №2(12) С. 27–35.
2. Жиликов Е.Г. Алгоритмы обнаружения основного тона речевых сигналов / Жиликов Е.Г., Фирсова А.А., Чеканов Н.А. — Научные ведомости БелГУ. Сер. История. Политология. Экономика. Информатика. — 2012. — №1–1(120), вып. 21. — С. 135–143.
3. Software for speech rehabilitation of cancer patients after larynx resection / S.S. Kharchenko, R.V. Meshcheryakov, D.A. Volf, L.N. Balatskaya, E.L. Choinzonov // Biomedical Engineering. — 2016. — Vol. 50, № 2. — P. 142–146.
4. Volf D. The singular estimation pitch tracker / D. Volf, R. Meshcheryakov, S. Kharchenko. — Proceedings of the 17th International Conference on Speech and Computer, SPECOM 2015. — 2015. — Athens, Greece. — P. 454–462.



5. Вишнякова О.А. Гибридный алгоритм выделения частоты основного тона / О.А. Вишнякова, Д.Н. Лавров. — Математические структуры и моделирование — 2015. — С. 59–65.
6. Schramm R. A polyphonic pitch tracking embedded system for rapid instrument augmentation / R. Schramm, F. Visi, A. Brasil, and M. Johann — NIME. — 2018. — Blacksburg, Virginia.
7. Yan Qi Copy-move detection of audio recording with pitch similarity / Qi Yan, Rui Yang, Jiwu Huang. — 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). — 2015.
8. Полешников Д.Д. Метод выделения изменения частоты сердечных сокращений из естественного речевого сигнала / Д.Д. Полешников, О.О. Басов. — Вестник воронежского института МВД России — 2019. — С. 55–63.
9. Войтенко В.В. Возможности спектрального анализа голоса для диагностики повреждения наружной ветви верхнего гортанного нерва в тиреоидной хирургии. // Клінічна ендокринологія та ендокринна хірургія. — 2015. — С. 46–50.
10. Харченко С.С. Математическая электроакустическая модель псевдоголоса и программный комплекс голосовой реабилитации пациентов после ларингэктомии на основе бионических принципов: дис. ... канд. техн. наук. Томск, 2017. — 114 с.
11. Алимуратов А.К. Оценка частоты основного тона речевых сигналов методами декомпозиции на эмпирические моды. // Измерение. Мониторинг. Управление. Контроль. — 2015. — С. 37–45.
12. Alimuradov A. Improved CEEMDAN Based Speech Signal Analysis Algorithm for Mental Disorders Diagnostic System: Pitch Frequency Detection and Measurement. / A. Alimuradov, A. Tychkov. // International Journal of Embedded and Real-Time Communication Systems (IJERTCS). — 2019. — P. 22–47.
13. Low-Latency Audio Pitch Tracking: a Multi-Modal Sensor-Assisted Approach. / S. Laurel et al. // Proceedings of the 2014 International Conference on New Interfaces for Musical Expression. — London. — UK. — P. 54–59.
14. Real-time pitch tracking in audio signals with the extended complex Kalman filter. / Orchisama Das et al. // Proceedings of the 20th International Conference on Digital Audio Effects. — 2017. — Edinburgh, UK.
15. Stone S. A Time-Warping Pitch Tracking Algorithm considering fast f0 changes. / S. Simon, P. Steiner, P. Birkholz. // INTERSPEECH. — 2011. — Stockholm, Sweden.
16. Drugman T. Joint Robust Voicing Detection and Pitch Estimation Based on Residual Harmonics. / T. Drugman, A. Alwan. // INTERSPEECH. — 2011. — Florence, Italy.
17. Han K. Neural Network Based Pitch Tracking in Very Noisy Speech. / K. Han, D. Wang. // IEEE/ACM Transactions on Audio, Speech, and Language Processing. — 2014. — P. 2158–2168.
18. Abhijith M.N. Multi-pitch tracking using Gaussian mixture model with time varying parameters and Grating Compression Transform. / M.N. Abhijith, P.K. Ghosh, K. Rajgopal. // IEEE/ACM Transactions on Audio, Speech, and Language Processing. — 2014. — Florence, Italy.
19. Crepe: A Convolutional Representation for Pitch Estimation. / Jong Wook Kim et al. // IEEE/ACM Transactions on Audio, Speech, and Language Processing. — 2018. — Calgary, AB, Canada.
20. Han K. Neural networks for supervised pitch tracking in noise. / K. Han, D. Wang. // IEEE/ACM Transactions on Audio, Speech, and Language Processing. — 2014. — Florence, Italy.
21. Su H. Convolutional neural network for robust pitch determination. / H. Su, H. Zhang. // IEEE/ACM Transactions on Audio, Speech, and Language Processing. — 2016. — Shanghai, China.

**Харченко С.С., Крючков И.А., Мещеряков Р.В.**

**Определение частоты основного тона речевого сигнала на основе индекса многомерной синхронизации**

22. Liu Y. Time and frequency domain long short-term memory for noise robust pitch tracking. / Y. Liu, D. Wang. // IEEE/ACM Transactions on Audio, Speech, and Language Processing. — 2017. — New Orleans, LA, USA.
23. Wang K. A Pitch-aware Approach to Single-channel Speech Separation. / K. Wang, F. Soong, L. Xie. // IEEE/ACM Transactions on Audio, Speech, and Language Processing. — 2019. — Brighton, United Kingdom, United Kingdom.
24. Mandarin tone classification without pitch tracking. // IEEE/ACM Transactions on Audio, Speech, and Language Processing. — 2014. — Florence, Italy.
25. Shi L. Robust Bayesian Pitch Tracking Based on the Harmonic Model. / L. Shi, J.K. Nielsen. // IEEE/ACM Transactions on Audio, Speech, and Language Processing. — 2019. — P. 1737–1751.
26. Berg E. Dictionary-based pitch tracking with dynamic programming. / E. van den Berg, B. Ramabhadran. // INTERSPEECH. — 2014. — P. 1347–1350.
27. Вольф Д.А. Модель и программная реализация сингулярного оценивания частоты основного тона речевого сигнала. / Д.А. Вольф, Р.В. Мещеряков. // Труды СПИИРАН. — 2015. — С. 191–209.
28. Петровский А.А. Обработка речевых сигналов в приложениях мультимедиа на основе периодической модели с нестационарными параметрами. / А.А. Петровский, И.С. Азаров. // Доклады Белорусского государственного университета информатики и радиоэлектроники. — 2014.
29. Detection of neural activity in functional MRI using canonical correlation analysis / O. Friman, J. Cedefamn, P. Lundberg, M. Borga and H. Knutsson // Magn Reson Med — 2001 — 45:323–30.
30. Алгоритмы обнаружения основного тона речевых сигналов. / Е.Г. Жилияков, А.А. Фирсова, Н.А. Чеканов. // Научные ведомости. — 2012. — С. 135–143.
31. Restrictions of the Measurement System and a Patient When Using Visually Evoked Potentials / Sergei Kharchenko, Yaroslav Turovsky, Roman Mescheryakov, Anastasia Iskhakova. // 12th International Conference on Developments in eSystems Engineering (DeSE). — 2019. — Kazan, Russia.
32. Implementation of Robot–Human Control Bio-Interface When Highlighting Visual-Evoked Potentials Based on Multivariate Synchronization Index / Sergei Kharchenko, Yaroslav Turovsky, Roman Mescheryakov, Daniyar Volf. // Proceedings of 15th International Conference on Electromechanics and Robotics «Zavalishin’s Readings». — 2020. — Ufa, Russia.
33. A Pitch Tracking Corpus with Evaluation on Multipitch Tracking Scenario. / G. Pirker, M. Wohlmayr, S. Petrik and F. Pernkopf // Proceedings of INTERSPEECH. — 2011. — p. 1509–1512.
34. Азаров И.С. Алгоритм оценки мгновенной частоты основного тона речевого сигнала. / И.С. Азаров, М.И. Вашкевич, А.А. Петровский. // Цифровая Обработка Сигналов. — 2012. — С. 49–57.

## **ADAPTIVE PRIMARY CODES OF SPEECH SIGNALS BASED ON THEOREM V.A. KOTELNIKOV AND REPRESENTATIONS OF KHURGIN-YAKOVLEV**

*Kharchenko S.S., PhD of Technical Sciences, Associate Professor of the Department of Information Systems Security, Tomsk State University of Control Systems and Radioelectronics, Tomsk, Russia, kss@fb.tusur.ru*

*Kryuchkov I.A., software developer of the component development department of JSC «InfoTeKS», Tomsk, Russia*



*Meshcheryakov R. V., PhD of Technical Sciences, Professor,  
Chief Researcher of the Laboratory of Cyber-physical Systems  
of the Institute of Control Sciences of the Russian Academy of  
Sciences, Moscow, Russia*

The article is devoted to the study of determining the frequency of the fundamental frequency of a speech signal. The material for the study was the database "The Pitch-Tracking Database from Graz University of Technology". The developed algorithm for determining the fundamental frequency of a speech signal based on the multidimensional synchronization index was used in the work. To interpret the results obtained, a comparison with known estimators was used; the following metrics were used: the percentage of gross errors and the average percentage of small errors. In the course of the study, it was found that the algorithm is more appropriate to use to determine the average pitch frequency than to determine the pitch frequency in real time.

• *fundamental frequency* • *multivariate synchronization index* • *speech signal* • *pitch tracking*.

# Интеллектуальный анализ и моделирование развития ситуаций на основе сочетания нейросетевой обработки текстов, методов нечёткого когнитивного анализа и динамической кластеризации

*Борисов В.В., доктор технических наук, профессор,  
Филиал НИУ «МЭИ» в г. Смоленске, профессор,  
г. Смоленск, Россия, vbor67@mail.ru*

*Харламов А.А., доктор технических наук, старший научный  
сотрудник ИВНД и НФ РАН, профессор МГЛУ, профессор ВШЭ,  
заведующий кафедрой МФТИ,  
Москва, Россия, kharlamov@analyst.ru*

Предложен и реализован оригинальный подход к интеллектуальному анализу и моделированию развития ситуаций на основе сочетания нейросетевой обработки текстов, методов нечеткого когнитивного анализа и динамической кластеризации. Искусственные нейронные сети используются для первичной обработки корпусов текстов, с выделением семантической сети концептов, характеризующих анализируемые ситуации. Использование нечеткого когнитивного подхода позволяет представить взаимосвязи и взаимозависимости между объектами ситуаций в виде нечетких отношений взаимовлияния, результаты анализа которых используются для обоснования совокупности показателей для идентификации и мониторинга динамики изменения кластерной структуры ситуаций, а также для анализа устойчивости идентифицированных групп ситуаций. Методы же динамической кластеризации применяются для мониторинга динамики изменения групп ситуаций, прогноза их развития в зависимости от специфических внешних воздействий, представленных текстовыми описаниями.

• анализ ситуаций • нейросетевая обработка текстов • однородная семантическая сеть • нечёткая когнитивная модель • динамическая кластеризация.

## ВВЕДЕНИЕ

Большие объемы информации, которые, с одной стороны, необходимы для успешного анализа, а с другой, возникают в настоящее время как реакция на текущие события

разной этиологии, предполагают исключительно автоматические механизмы анализа этой информации.

Информация в текстах характеризует описанные в них ситуации как в статике, так и в динамике. Эта информация характеризуется различной степенью структурированности, многоаспектностью, лингвистической неопределенностью, многозначностью и противоречивостью, неполнотой, а также многообразием и разнотипностью взаимосвязей и взаимозависимостей между концептами анализируемой проблемной области. Текст, представленный (с использованием искусственных нейронных сетей [1]) в виде однородной семантической сети, является смысловым портретом анализируемой ситуации. Группы ситуаций структурно определяются устойчивыми кластерами взаимосвязанных и взаимозависимых пар концептов предметной области. Такое представление позволяет выявлять различные ситуации, оценивать их устойчивость, сопоставлять их между собой, моделировать динамику их структуры, осуществлять мониторинг и прогнозировать развитие ситуаций. Однако отмеченная выше специфика анализируемой информации существенно затрудняет использование типовых подходов к кластеризации концептов с целью идентификации групп ситуаций.

Использование нечеткого когнитивного подхода обеспечивает представление взаимосвязей и взаимозависимостей между концептами (объектами) ситуаций в виде нечетких отношений взаимовлияния, и, таким образом, обеспечивая в последующем возможность использования и развития исследовательского инструментария нечеткой динамической кластеризации [2] для анализа и моделирования развития ситуаций.

Выделение взаимосвязанных кластеров семантической сети, состоящих из выделенных в тексте концептов, описывающих ситуацию с определенной точки зрения, позволяет анализировать сущности (информационные объекты) ситуации в их коллаборации/противопоставлении, а также устойчивость ситуации в динамике, прогнозировать развитие самых разнообразных процессов, в том числе социально-экономических, социотехнических и других.

В дальнейшем для каждого идентифицированного класса ситуаций может быть построена своя модель, предназначенная для решения спектра задач интеллектуального анализа, прогнозирования и диагностирования ситуаций.

В процессе «эволюции» проблемной области неизбежно изменяются взаимосвязи и взаимозависимости между выявленными концептами, что приводит к изменению самой кластерной структуры этих пар концептов, ее устойчивости и, как следствие, к перегруппировке ситуаций.

Для мониторинга динамики изменения классов ситуаций применяются методы динамической кластеризации. Мониторинг заключается в анализе изменений кластерной структуры пар концептов, включая дрейф

центров кластеров, исчезновение и появление новых кластеров, их объединение и разделение.

## **1. ПЕРВИЧНАЯ ОБРАБОТКА ТЕКСТОВОГО КОРПУСА, ОПИСЫВАЮЩЕГО СИТУАЦИЮ**

Для нечеткой кластеризации по основанию «степень воздействия» требуется выбор из текста направленных пар концептов, из которых один (первый в паре) воздействует на другой с силой, представленной в тексте.

В качестве инструментария для выявления таких пар концептов в текстах используется нейросетевая технология TextAnalyst для автоматического смыслового анализа текстов [1]. Функциональность технологии позволяет на основе анализируемого корпуса текстов построить семантическую сеть корпуса, вершинами которой являются концепты (объекты), представленные в текстах, а дуги характеризуют связи этих концептов в тексте. И концепты, и их связи маркируются их смысловыми весами. Весовые характеристики концептов характеризуют их важность в тексте, которая выявляется с учетом связности конкретного концепта с другими концептами на несколько шагов по сети. Весовые характеристики связей характеризуют степень связности отдельных концептов в тексте.

И сеть, и весовые характеристики ее вершин и связей формируются автоматически. Формирование семантической сети проводится в два этапа [3]. На первом этапе осуществляется формирование частотной сети на основе выявленных в корпусе текстов частоты встречаемости слов в тексте и встречаемости пар слов в предложениях текста. На втором этапе осуществляется переранжирование весовых характеристик концептов с помощью итеративной процедуры, учитывающей вклад связанных в сети с концептами понятий на несколько шагов.

### **1.1. Технология для анализа текстов**

#### **1.1.1. Первичная обработка текста**

Обычно текст содержит кроме текстовой также нетекстовую информацию: рисунки, таблицы, графики, адреса электронной почты, сокращения и аббревиатуры. Кроме того, для построения однородной семантической сети необходимо убрать из текста также рабочие слова (предлоги, артикли), общеупотребимые слова. И еще необходимо текст разбить на слова, предложения, а слова подвергнуть процедуре стемминга, чтобы убрать влияние словоизменения на результаты анализа.

#### **1.1.2. Формирование первичной сети**

В отличие от подхода, основанного на дистрибутивной семантике (где текст рассматривается как мешок слов), для анализа текста формируется сеть, вершинами которой являются понятия, представленные словами текста, а дуги соответствуют отношениям этих понятий в действительности, описанной текстом. Такая сеть позволяет рассматривать связи слов в тексте на любую глубину (как цепи вершин), а потому учитывать влияние предшествующих слов на последующие.





Первичная сеть строится на основе анализа словоупотребления в анализируемом тексте, а также употребления пар слов в предложениях текста. Первоначальные веса  $z_i$  вершин сети оцениваются частотой встречаемости слова в тексте. Веса дуг  $z_{ij}$  — частотой встречаемости пар слов в предложениях текста.

**Семантическая сеть.** Формально ассоциативная (однородная семантическая) сеть  $SN$  может быть представлена следующим образом [4].

**Определение 1.** Под семантической сетью  $SN$  понимается множество несимметричных пар событий  $\{ \langle c_i, c_j \rangle \}$ , где  $c_i$  и  $c_j$  — события, связанные между собой отношением ассоциативности (совместной встречаемости в некоторой ситуации):

$$SN \cong \{ \langle c_i, c_j \rangle \}$$

Применительно к содержанию статьи этими событиями являются слова  $w_i$  текста. Следующие друг за другом пары слов  $(w_i, w_j)$  текста и составляют однородную семантическую (ассоциативную) сеть:

$$SN \cong \{ \langle w_i, w_j \rangle \}$$

В данном случае отношение ассоциативности несимметрично:

$$\langle w_i, w_j \rangle \neq \langle w_j, w_i \rangle.$$

**Определение 2.** Семантическая сеть, описанная таким образом, может быть представлена как множество так называемых звездочек  $\{ \langle w_i \langle w_j \rangle \rangle \}$ :

$$SN \cong \{ z_i \} = \{ \langle w_i \langle w_j \rangle \rangle \}$$

**Определение 3.** Имея в виду, что события  $w_i$  и  $w_j$  есть слова, под звездочкой  $\langle w_i \langle w_j \rangle \rangle$  понимается конструкция, включающая главное слово  $w_i$ , связанное с множеством слов-ассоциантов  $\{ w_j \}$ , которые являются семантическими признаками главного слова, отстоящими от главного слова в сети на одну связь. Связи направлены от главного слова к словам-ассоциантам.

### 1.1.3. Ранжирование вершин и дуг

Поскольку реально ранг вершин отличается от частоты встречаемости слов в тексте (на ранг вершины влияют веса вершин, связанных с ней на один и более шагов на сети), для расчета ранга вершин используется итеративная процедура переранжирования, где суммарный нормированный вес вершин, связанных с анализируемой вершиной, влияет на ее ранг, а процедура проводится до сходимости процесса.

В отличие от веса вершин, веса дуг остаются равными первоначально посчитанной частоте  $z_{ij}$  встречаемости пар слов в предложениях текста. Она далее и используется для проведения нечеткой кластеризации. Необходимо помнить, что пары слов направленные, то есть влияние одного концепта на другой в общем случае не одинаково с влиянием второго концепта на первый.

Единственное изменение масштаба, которое необходимо для последующего анализа, это переход от шкалы  $[0 \div 100]$  к шкале  $[-50 \div +50]$ , так как для реализации нечеткой кластеризации требуется учет как корреляции, так и антикорреляции.

***N*-граммная модель текста.** При анализе текста используется *n*-граммная, а точнее, «односторонняя» *n*-граммная модель, а именно принятая при использовании *n*-грамм «правосторонняя» модель, в которой вероятность очередного слова строки задается в зависимости от предшествующих ему  $(n - 1)$  слов, что записывается как  $p(w_n | w_1, \dots, w_{n-1})$ . Тогда:

$$p(w_1, \dots, w_{n-1}w_n) = p(w_n | w_1, \dots, w_{n-1}) p(w_1 \dots w_{n-1}).$$

В терминах вероятности «быть справа» имеем:

$$p(w_1 \dots w_{n-1}w_n) = p(w_n | w_1 \dots w_{n-1})p(w_{n-1} | w_1 \dots w_{n-2}) \dots p(w_3 | w_1w_2)p(w_2 | w_1)p(w_1),$$

или

$$p(w_1 \dots w_n) = \left( \prod_{k=1}^n p(w_k | w_1 \dots w_{k-1}) \right) p(w_1),$$

Введя фиктивный символ «начало» и договорившись, что  $p(w_1 | w_0) \cong p(w_1)$ , можно переписать:

$$p(w_1 \dots w_n) = \left( \prod_{k=1}^n p(w_k | w_1 \dots w_{k-1}) \right) p(w_1),$$

Оценкой вероятности *n*-граммы служит частота ее встречаемости:

$$\hat{p}(w_i | w_{i-n} \dots w_{i-1}) = f(w_i | w_{i-n} \dots w_{i-1}) = \frac{C(w_{i-n} \dots w_{i-1}w_i)}{C(w_{i-n} \dots w_{i-1})}.$$

Так для биграммной модели оценкой вероятности биграммы является частота ее появления в тексте:

$$\hat{p}(w_j | w_i) = f(w_j | w_i) = \frac{C(w_jw_i)}{C(w_i)}.$$

Использование *n*-граммной модели хоть и более корректно при анализе текстов с лингвистической точки зрения, увы, невозможно при  $n > 3$  для текстов нефлективных языков, а для флективных языков *n* не может превысить 2, так как в мире не существует такого количества текстов соответствующего языка для обучения модели соответствующей граммности [5]. Для русского языка пользуются, например, 2,5-граммными моделями, где пары рассматривают как пары слов, а в тройки включают наряду с парой слов третьим элементом — грамматическую категорию вместо соответствующего слова. Либо используют другие эмпирические «натяжки».

Использование частного случая представления текста в виде однородной семантической сети позволяет ввести *n*-граммность аналитически [6] (на сети).

***Биграммная модель.*** Если мы будем рассматривать последовательности из двух слов на сети текста, получим биграммную модель. Для каждого второго слова  $w_j$  строки из двух слов  $(w_i, w_j)$  первое слово строки  $w_i$  («слева-направо»)



является темой:  $w_i \cong t_i^2$  (индекс «2» — поскольку биграммная модель). Объединим все пары слов с одинаковыми темами в звездочки. В этом случае вершина сети, соответствующая слову  $w_i \cong t_i^2$ , является корневой вершиной одного из тематических деревьев (в данном случае — звездочек). Так как вероятность появления строки из двух слов (правосторонняя модель) в тексте  $p(w_i, w_j) = p(w_j|w_i)p(w_i)$ , вероятность появления темы в биграммной модели есть сумма вероятностей появления пар с одинаковым первым словом (вероятность появления звездочки):

$$p(t_i) = p(w_i^2) = \sum_{j=1}^{J_i} p(w_i, w_j) = \sum_{j=1}^{J_i} p(w_j|w_i)p(w_i^1),$$

где  $J_i$  — число слов  $w_j$  в сети (ассоциантов  $w_j$  звездочки), связанных с первым словом  $w_i$ . А вероятность  $p(w_i^1)$  — это исходные вероятности распределения слов в тексте. Введем условное понятие «темы»  $p(w_i^1) \cong t_i^1$  для монограммного распределения. И так для каждого второго слова первое слово пары («слева-направо») является темой:  $w_i \cong t_i^2$ :

$$p(t_i^2) = p(w_i^2) = \sum_{j=1}^{J_i} p(w_i, w_j) = \sum_{j=1}^{J_i} p(w_j|w_i)p(t_i^1),$$

Для того чтобы общая полученная сумма вероятностей  $P(t_i^2)$  была равна единице:

$$\sum_{n=1}^N P(t_n^2) = 1,$$

где  $n$  — число тем, необходимо нормировать полученные для каждой темы  $t_n^1$  суммы. В общем случае число тем совпадает с числом всех слов текста  $T=W$ , но обычно выбирают лишь несколько главных тем:  $T \leq W$ .

Нормирование осуществляется на сумму по всем темам  $t_i^1$ :

$$p(t_i^2) = \frac{\sum_{j=1}^{J_i} p(w_j|w_i)p(t_i^1)}{\sum_{i=1}^T \sum_{j=1}^{J_n} p(w_j|w_i)p(t_i^1)}.$$

Здесь  $p(t_i^1)$  означает вероятность появления отдельного слова (то есть — вероятность из монограммной модели). А  $p(t_i^2)$  — вероятность появления звездочки из биграммной модели. Причём главное слово звездочки в терминах биграммной модели является темой для ее ближайших ассоциантов — семантических признаков — «вторых» слов.

**Триграммная модель.** Рассмотрим теперь не последовательности длиной в два слова, а последовательности длиной в три слова в сети текста, то есть триграммную модель:

$$p(w_i w_j w_k) = p(w_k|w_i w_j)p(w_i w_j) = p(w_k|w_i w_j)p(w_j|w_i)p(w_i).$$

Тогда вероятность появления строки из первых двух слов в строке из трех слов можно получить, как и в (1), суммированием по третьему слову:

$$p(w_i w_j) = \sum_{k=1}^{K_j} p(w_i w_j w_k),$$

и вероятность появления первого слова в строке из трех слов можно получить суммированием по второму слову:

$$p(w_i) = \sum_{j=1}^{J_i} \sum_{k=1}^{K_j} p(w_i w_j w_k).$$

Тогда  $p(t_i^3)$  вычисляется в рамках биграммной модели,  $p(w_i) \cong p(t_i^2)$ , а вероятность появления второго слова пары в зависимости от появления первого слова пары в тексте одна и та же, независимо от граммотности модели текста, как следует из сетевого представления текста (взаимосвязанные пары слов остаются теми же):

$$p(t_i^3) = \frac{\sum_{j=1}^{J_i} p(w_j | t_i^2) p(t_i^2)}{\sum_{i=1}^I \sum_{j=1}^{J_i} p(w_j | t_i^2) p(t_i^2)},$$

где  $t_i^2$  и  $t_i^3$  обозначают, соответственно, темы в соответствии с биграммной и триграммной моделями, а число тем, как и ранее, может соответствовать числу слов в тексте  $T=W$ , но обычно ограничивается волевым решением до  $T=W$ . К корневым вершинам  $t_i^3$  крепятся через «вторые» слова-вершины  $t_i^2$  звездочки  $Z_i^3$  звездочки биграммной модели  $Z_i^2$ :

$$Z_i^2 = \{ \langle t_i^2 \{ t_i^1 \} \rangle \}.$$

**Четырех- и более граммотные модели. Итеративная процедура перевзвешивания.**

Та же процедура, примененная для анализа четырехграммотной модели, приведет к получению следующего уровня множества корневых вершин, к которым крепятся все графы (тематические структуры), полученные на этапе анализа триграммотной модели. То есть мы можем вычислить веса тематических вершин семантической сети текста с учетом их зависимости в последовательности из четырех слов.

То же будет и для  $n$ -граммотной модели. Хотя надо заметить, что начиная с некоторого  $n$  процесс сойдется, так как зависимость слов, как правило, в модели языка сказывается не далее, чем на длине простого предложения.

Другими словами, мы имеем итеративную процедуру перевзвешивания, которая позволяет найти значения вероятностей появления тем  $t_i^n$  (в случае использования сетевой  $n$ -граммотной модели) в тексте.

$$p(t_i^2) = \frac{\sum_{j=1}^{J_i} p(w_j | w_i) p(t_i^1)}{\sum_{i=1}^I \sum_{j=1}^{J_n} p(w_j | w_i) p(t_i^1)},$$

$$p(t_i^3) = \frac{\sum_{j=1}^{J_i} p(w_j | w_i) p(t_i^2)}{\sum_{i=1}^I \sum_{j=1}^{J_i} p(w_j | w_i) p(t_i^2)},$$

...

$$p(t_i^n) = \frac{\sum_{j=1}^{J_i} p(w_j | w_i) p(t_i^{n-1})}{\sum_{i=1}^I \sum_{j=1}^{J_i} p(w_j | w_i) p(t_i^{n-1})},$$

где  $p(t_i^1) = p(w_2)$  и  $p(w_j | w_i)$  — одинаковая для всех шагов итерации вероятность появления последующего слова текста при условии появления предыдущего слова.

## 1.2. Текстовая информация для анализа

В качестве исходного материала для анализа ситуаций рассматривается корпус текстов, описывающих ситуацию, например выборка сообщений из социальной сети — откликов участников сети на некоторое событие [7].

Поскольку корпус текстов содержит самую различную информацию, необходимо предварительное выявление концептов, интересных с точки зрения последующего анализа — с точки зрения последующих возможных участников ситуации. Этот анализ проводится экспертами, которые формируют перечень концептов, используемых для дальнейшего анализа. Тем не менее и другая, не относящаяся к задаче информация оказывается важной при выявлении весовых характеристик семантической сети, так как она является контекстом, в котором реализуется анализируемая ситуация.

Выявленные экспертами концепты далее анализируются с точки зрения их весомости в корпусе текстов (в анализируемой ситуации): только концепты, ранги которых превышают пороговое значение, участвуют в последующем анализе. Для отобранных таким образом концептов исследуются далее весовые характеристики их положительных и отрицательных связей.

## 2. АНАЛИЗ НЕЧЕТКИХ ОТНОШЕНИЙ ВЗАИМОВЛИЯНИЯ МЕЖДУ КОНЦЕПТАМИ И ИДЕНТИФИКАЦИЯ КЛАСТЕРОВ (ГРУПП) СИТУАЦИЙ

Для выбранных концептов проблемной области  $A = \{a_i | i=1, \dots, I\}$  на основе их взаимосвязей задаются нечеткие отношения взаимовлияния:

$$R = \{(r_{ij} / (a_p a_j)) \mid r_{ij} \in [-1, 1], a_p a_j \in A\}.$$

Затем элементы отношения  $R$  преобразуются в элементы отношения  $Q$  по правилам:

$$\begin{aligned} \text{Если } r_{ij} > 0, \text{ то } q_{2i-1, 2j-1} &= r_{ij} \text{ и } q_{2i-1, 2j} = r_{ij}, \\ \text{Если } r_{ij} < 0, \text{ то } q_{2i-1, 2j-1} &= -r_{ij} \text{ и } q_{2i-1, 2j} = -r_{ij}. \end{aligned}$$

После чего выполняется транзитивное замыкание отношения  $Q$  основе операций max-prod-композиции:

$$\tilde{Q} = Q \vee Q^2 \vee Q^3 \vee \dots$$

Затем выполняется обратное преобразование отношения  $\tilde{Q}$  в отношение  $R' = \{(r'_{ij}, \bar{r}'_{ij})\}$ , состоящее из положительно-отрицательных пар согласованных элементов:

$$r'_{ij} = \max(q_{2i-1, 2j-1}, q_{2i, 2j}), \bar{r}'_{ij} = -\max(q_{2i-1, 2j}, q_{2i, 2j-1}).$$

В итоге рассчитываются следующие показатели для кластеризации концептов проблемной области [8]:

$$\text{воздействие объекта на ситуацию, } p_1(a_i) = \frac{1}{I} \sum_{j=1}^I (\text{sign}(r'_{ij} + \bar{r}'_{ij}) \max(|r'_{ij}|, |\bar{r}'_{ij}|)), \quad i = 1..I;$$

$$\text{воздействие ситуации на объект, } p_2(a_i) = \frac{1}{I} \sum_{j=1}^I (\text{sign}(r'_{ji} + \bar{r}'_{ji}) \max(|r'_{ji}|, |\bar{r}'_{ji}|)), \quad i = 1..I;$$

$$\text{согласованное влияние объекта на ситуацию, } p_3(a_i) = \frac{1}{I} \sum_{j=1}^I \frac{|r'_{ij} + \bar{r}'_{ij}|}{|r'_{ij}| + |\bar{r}'_{ij}|}, \quad i = 1..I.$$

На основе этих показателей и выполняется кластеризация концептов проблемной области и идентификация групп ситуаций.

### 3. МОДЕЛИРОВАНИЕ РАЗВИТИЯ СИТУАЦИЙ

После первоначальной идентификации групп ситуаций для мониторинга динамики их изменения целесообразно воспользоваться одним из методов нечеткой кластеризации, позволяющим определять нечеткие степени принадлежности объектов к кластерам.

Использование метода нечеткой кластеризации *fuzzy C-means* зачастую приводит к некорректным результатам в случаях, когда кластеры различаются по форме [9]. Этого недостатка лишен метод UOFC, представляющий собой комбинацию метода *fuzzy C-means* и усовершенствованного метода максимального правдоподобия [10]. Важным является и то, что данный метод позволяет найти наилучшее (в смысле показателей качества кластеризации) разбиение на кластеры.

Особенностью постановки и решения задачи мониторинга динамики изменения групп ситуаций на основе метода UOFC является то, что источником информации служит изменение значений нечетких отношений взаимовлияния между объектами.

Представленные выше процедуры осуществляются на корпусах текстов для разных временных промежутков («позавчера-вчера-сегодня»). Для каждого временного промежутка строится своя семантическая сеть, каждая из которых характеризует ситуации в этот промежуток времени.

При этом сам мониторинг заключается в анализе в каждый момент модельного времени: во-первых, изменений кластерной структуры проблемной области, включая



Таблица 1

Концепты		$p_1(a_i)$	$p_2(a_i)$	$p_3(a_i)$	$C(I)$
$a_1$	Авиасообщение	0.060	-0.098	-0.014	c3
$a_2$	Больницы	0.056	-0.162	-0.121	c4
$a_3$	Вузы	0.062	-0.013	0.193	c2
$a_4$	Выявлено	0.065	-0.063	0.033	c3
$a_5$	Гостиницы	0.042	0.196	-0.134	c1
$a_6$	Дезинфекция	0.049	0.182	-0.032	c1
$a_7$	Избегать	0.041	-0.025	0.006	c6
$a_8$	Италия	0.047	0.032	0.151	c2
$a_9$	Карантин	0.062	-0.126	0.017	c3
$a_{10}$	Китай	0.039	-0.028	0.151	c2
$a_{11}$	Коммуналка	0.041	0.088	0.061	c5
$a_{12}$	Коронавирусная	0.039	0.043	-0.304	c7
$a_{13}$	Магазины	0.038	0.123	0.098	c5
$a_{14}$	Медицинские работники	0.057	-0.074	-0.137	c4
$a_{15}$	Медицинские учреждения	0.051	-0.013	0.094	c2
$a_{16}$	Москва	0.057	0.006	0.014	c6
$a_{17}$	Нерабочая	0.035	0.003	0.209	c2
$a_{18}$	Ограничительные меры	0.058	0.080	-0.014	c7
$a_{19}$	Пенсионеры	0.026	0.036	0.031	c6
$a_{20}$	Регионы	0.038	0.020	-0.015	c6
$a_{21}$	Рестораны	0.033	0.074	0.102	c5
$a_{22}$	Симптомы	0.056	-0.032	0.086	c6
$a_{23}$	Случаи заражения	0.051	0.071	-0.079	c7
$a_{24}$	Средства индивидуальной защиты	0.059	0.142	-0.120	c1
$a_{25}$	Транспорт	0.036	0.138	0.130	c5
$a_{26}$	Ухань	0.042	0.022	-0.098	c7
$a_{27}$	Школы	0.039	-0.161	0.149	c3

анализ дрейфа центров кластеров (групп) ситуаций, исчезновения и появления новых кластеров, их объединения и разделения; во-вторых, устойчивости кластеров на основе оценки результатов транзитивного замыкания нечетких отношений взаимовлияния между объектами проблемной области.

#### 4. ПРИМЕР

В качестве примера рассмотрим задачу идентификации и моделирования развития ковид-ситуаций на начальной стадии, характеризующейся высокой динамикой изменения проблемной области за период с 26 по 31 марта 2019 г., разбитые на три временных интервала (эпохи): 26–27, 28–29 и 30–31 марта.

В результате анализа корпуса текстов общим объемом 180 Мб с использованием нейросетевой технологии *TextAnalyst* построена семантическая сеть, включающая в себя около 5 тыс. понятий (концептов) проблемной области.

Экспертно-статистическим методом отобрано 27 концептов, в наибольшей степени характеризующих ситуации рассматриваемой проблемной области, построена ее нечеткая когнитивная модель, и на основе согласованных нечетких отношений взаимовлияния между ее концептами рассчитаны значения показателей для идентификации кластеров этой проблемной области в первую эпоху (см. табл. 1).

На рис. 1 показано первоначальное расположение концептов и центров кластеров (идентифицированных групп ситуаций) в первую эпоху.

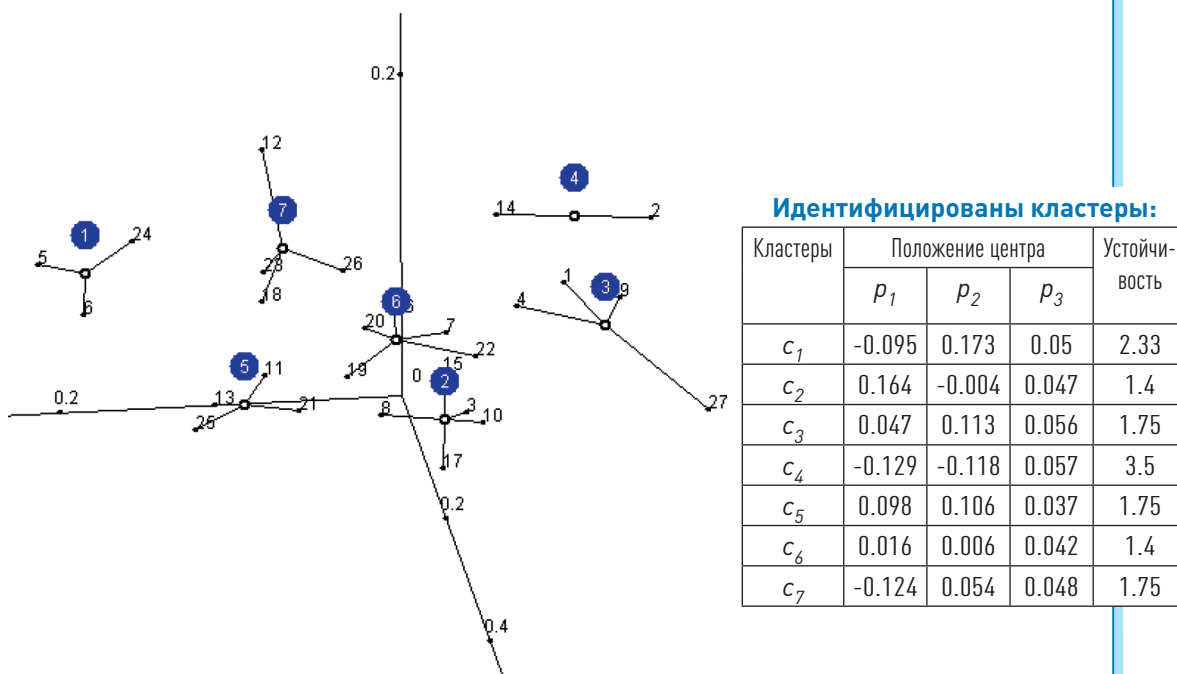


Рис. 1. Результаты первоначальной кластеризации концептов проблемной области в первую эпоху

Транзитивное замыкание нечеткой когнитивной модели свидетельствует об устойчивости кластерной структуры проблемной области в целом. Такой же вывод справедлив и для всех идентифицированных кластеров (семь групп ситуаций).

На рис. 2 и 3 отражено изменение кластерной структуры проблемной области во вторую и третью эпохи.





**Остались кластеры (группы ситуаций):**

Кластеры	Положение центра			Устойчивость
	$p_1$	$p_2$	$p_3$	
$c_1$	0.032	0.173	0.05	0.86
$c_2$	0.164	-0.004	0.047	0.67
$c_5$	0.047	0.113	0.056	0.67
$c_6$	-0.129	-0.118	0.057	3.0

Исчезли кластеры:  $c_3, c_4, c_7$ .

Переходы концептов:

- $a_1 : c_3 \rightarrow c_1$
- $a_2 : c_4 \rightarrow c_5$
- $a_3 : c_2 \rightarrow c_5$
- $a_4 : c_3 \rightarrow c_5$
- $a_6 : c_1 \rightarrow c_2$
- $a_7 : c_6 \rightarrow c_2$
- $a_9 : c_3 \rightarrow c_5$
- $a_{12} : c_7 \rightarrow c_5$
- $a_{13} : c_5 \rightarrow c_2$
- $a_{14} : c_4 \rightarrow c_1$
- $a_{18} : c_7 \rightarrow c_5$
- $a_{19} : c_6 \rightarrow c_2$
- $a_{22} : c_6 \rightarrow c_1$
- $a_{23} : c_7 \rightarrow c_1$
- $a_{26} : c_7 \rightarrow c_1$
- $a_{27} : c_3 \rightarrow c_2$

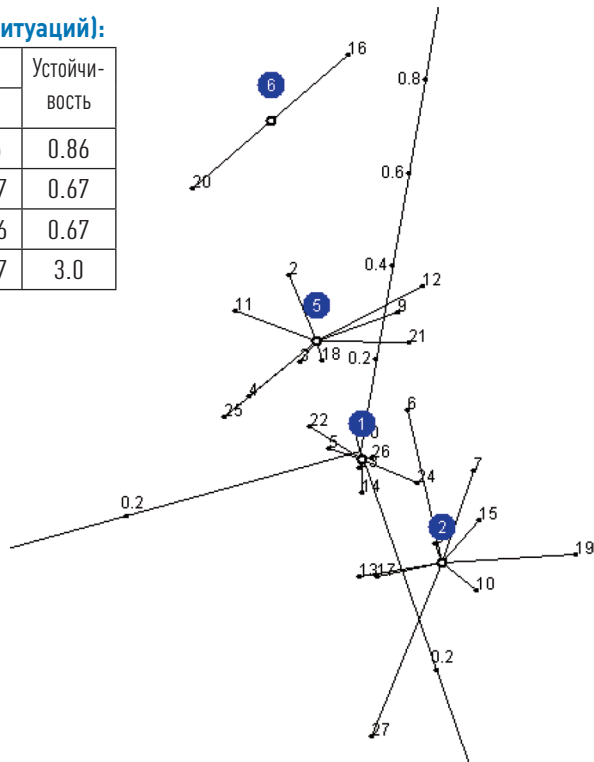


Рис. 2. Результаты кластеризации концептов проблемной области во вторую эпоху

**Остались кластеры (группы ситуаций):**

Кластеры	Положение центра			Устойчивость
	$p_1$	$p_2$	$p_3$	
$c_1$	0.032	0.173	0.05	1.5
$c_2$	0.164	-0.004	0.047	1.0
$c_5$	0.047	0.113	0.056	3.0
$c_6$	-0.129	-0.118	0.057	3.0
$c_8$	0.099	0.116	0.064	3.0

Появился кластер  $c_8$ .

Переходы концептов:

- $a_3 : c_5 \rightarrow c_8$
- $a_4 : c_5 \rightarrow c_2$
- $a_6 : c_2 \rightarrow c_1$
- $a_8 : c_2 \rightarrow c_8$
- $a_9 : c_5 \rightarrow c_6$
- $a_{10} : c_2 \rightarrow c_8$
- $a_{12} : c_5 \rightarrow c_6$
- $a_{14} : c_1 \rightarrow c_2$
- $a_{16} : c_6 \rightarrow c_2$
- $a_{17} : c_2 \rightarrow c_5$
- $a_{18} : c_5 \rightarrow c_1$
- $a_{19} : c_2 \rightarrow c_8$
- $a_{21} : c_5 \rightarrow c_8$
- $a_{23} : c_1 \rightarrow c_8$
- $a_{24} : c_1 \rightarrow c_2$
- $a_{25} : c_5 \rightarrow c_2$
- $a_{26} : c_1 \rightarrow c_2$
- $a_{27} : c_2 \rightarrow c_1$

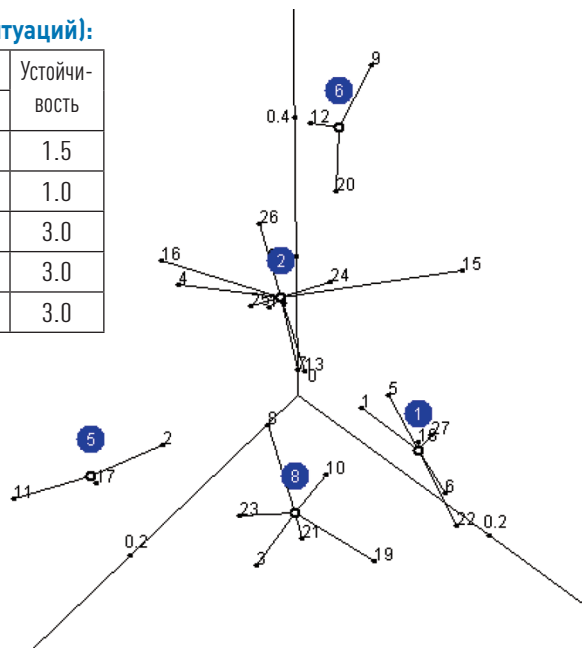


Рис. 3. Результаты кластеризации концептов проблемной области в третью эпоху

В табл. 2 отражена динамика изменения кластерной структуры проблемной области за три эпохи.

Таблица 2

Концепты		$C_{(I)}$	$C_{(II)}$	$C_{(III)}$
$a_1$	Авиасообщение	$c_3$	$c_1$	$c_1$
$a_2$	Больницы	$c_4$	$c_5$	$c_5$
$a_3$	Вузы	$c_2$	$c_5$	$c_8$
$a_4$	Выявлено	$c_3$	$c_5$	$c_2$
$a_5$	Гостиницы	$c_1$	$c_1$	$c_1$
$a_6$	Дезинфекция	$c_1$	$c_2$	$c_1$
$a_7$	Избегать	$c_6$	$c_2$	$c_2$
$a_8$	Италия	$c_2$	$c_2$	$c_1$
$a_9$	Карантин	$c_3$	$c_5$	$c_6$
$a_{10}$	Китай	$c_2$	$c_2$	$c_8$
$a_{11}$	Коммуналка	$c_5$	$c_5$	$c_5$
$a_{12}$	Коронавирусная	$c_7$	$c_5$	$c_6$
$a_{13}$	Магазины	$c_5$	$c_2$	$c_2$
$a_{14}$	Медицинские работники	$c_4$	$c_1$	$c_2$
$a_{15}$	Медицинские учреждения	$c_2$	$c_2$	$c_2$
$a_{16}$	Москва	$c_6$	$c_6$	$c_2$
$a_{17}$	Нерабочая	$c_2$	$c_2$	$c_5$
$a_{18}$	Ограничительные меры	$c_7$	$c_5$	$c_1$
$a_{19}$	Пенсионеры	$c_6$	$c_2$	$c_8$
$a_{20}$	Регионы	$c_6$	$c_6$	$c_6$
$a_{21}$	Рестораны	$c_5$	$c_5$	$c_8$
$a_{22}$	Симптомы	$c_6$	$c_1$	$c_1$
$a_{23}$	Случаи заражения	$c_7$	$c_1$	$c_8$
$a_{24}$	Средства индивидуальной защиты	$c_1$	$c_1$	$c_2$
$a_{25}$	Транспорт	$c_5$	$c_5$	$c_2$
$a_{26}$	Ухань	$c_7$	$c_1$	$c_2$
$a_{27}$	Школы	$c_3$	$c_2$	$c_1$

## ЗАКЛЮЧЕНИЕ

Рассмотрено решение комплексной задачи идентификации и моделирования развития ситуаций проблемной области на основе нейросетевой обработки текстов и сочетания методов нечеткого когнитивного анализа и динамической кластеризации.

Интеллектуальная нейросетевая система/технология используется для предварительного анализа проблемной области, заключающегося в обработке разнородной текстовой информации, анализе, выявлении и оценке значимости объектов проблемной области, характеризующих группы ситуаций.

Нечеткий когнитивный подход создает основу для идентификации групп ситуаций за счет представления взаимосвязей и взаимозависимостей между выявленными концептами ситуаций в виде нечетких отношений взаимовлияния, обеспечивая возможность обоснования показателей кластеризации объектов.

Для мониторинга динамики изменения групп ситуаций, т. е. изменения кластерной структуры объектов проблемной области, включая дрейф центров кластеров (групп) ситуаций, исчезновение и появление новых кластеров, их объединение и разделение, применяются методы динамической кластеризации.

Для каждой идентифицированной группы ситуаций в каждой эпохе может быть построена своя модель для интеллектуального анализа, прогнозирования и диагностирования ситуаций.

В качестве примера рассмотрена задача идентификации и моделирования развития ковид-ситуаций на начальной стадии, характеризующейся высокой динамикой изменения проблемной области.

*Исследование выполнено при финансовой поддержке Министерства науки и высшего образования РФ (государственное задание № FSWF-2020-0019).*

### Литература

1. Харламов А.А. Ассоциативная память — среда для формирования пространства знаний. От биологии к приложениям. — Дюссельдорф: Palmarium Academic Publishing, 2017.
2. Angstenberger L. Dynamic fuzzy pattern recognition with applications to finance and engineering. Boston: Kluwer Academic Publishers, 2001.
3. Харламов А.А. Нейросетевая технология представления и обработки информации (естественное представление знаний). — М.: Радиотехника, 2006.
4. Харламов А.А., Ермоленко Т.В. Нейросетевая среда (нейроморфная ассоциативная память) для преодоления информационной сложности. Поиск смысла в слабоструктурированных массивах информации. Часть II. Обработка информации в гиппокампе. Модель мира / Информационные технологии, N 12, 2015. — Стр. 883—889.
5. Воронцов К.В. Вероятностные тематические модели (курс лекций).
6. Харламов А.А. Формирование n-граммной тематической модели текста// Речевые технологии № 1–2, 2016. С. 15–23.
7. Kharlamov A.A., Raskhodchikov A.N., Pilgun M. Smart City Data Sensing during COVID-19: Public Reaction to Accelerating Digital transformation. Sensors 2021, 21(12), 3965

**Борисов В.В., Харламов А.А.**

**Интеллектуальный анализ и моделирование развития ситуаций на основе сочетания нейросетевой обработки текстов, методов нечеткого когнитивного анализа и динамической кластеризации**

8. Арбузов А.Д., Борисов В.В., Дли М.И. Мониторинг кластеров социотехнических систем на основе нечеткого когнитивного подхода и методов динамической кластеризации // Мягкие измерения и вычисления. 2020. Т. 33. № 8. — С. 26–37.
9. Bezdek J.C., Keller J., Krisnapuram R., Pal N. Fuzzy Models and Algorithms for Pattern Recognition and Image Processing. New York, Springer Science, 2005.
10. Geva A.B., Steinberg Y., Bruckmair S., Nahum G. A Comparison of Cluster Validity Criteria for a Mixture of Normal Distributed Data // Pattern Recognition Letters. 2000. Vol. 21(6–7), pp. 511–529.

## **INTELLIGENT ANALYSIS AND MODELING OF THE DEVELOPMENT OF SITUATIONS BASED ON A COMBINATION OF NEURAL NETWORK PROCESSING OF TEXTS, METHODS OF FUZZY COGNITIVE ANALYSIS AND DYNAMIC CLUSTERING**

***Borisov V.V., Doctor of Technical Sciences, Professor,  
Branch of NRU "MPEI" in Smolensk, Professor,  
Smolensk, Russia, vbor67@mail.ru***

***Kharlamov A.A., Doctor of Technical Sciences, Senior Researcher  
employee of IHNA and NP RAS, Moscow, professor of Moscow State  
Linguistic University, HSE Professor, Head of Department,  
Moscow Institute of Physics and Technology,  
Moscow, Russia, kharlamov@analyst.ru***

An original approach to the intellectual analysis and modeling of the development of situations based on a combination of neural network text processing, fuzzy cognitive analysis methods and dynamic clustering is proposed and implemented. Artificial neural networks are used for the primary processing of text corpora, with the selection of a semantic network of concepts that characterize the analyzed situations. The use of a fuzzy cognitive approach makes it possible to represent the relationships and interdependencies between the objects of situations in the form of fuzzy relations of mutual influence, the results of the analysis of which are used to substantiate a set of indicators for identifying and monitoring the dynamics of changes in the cluster structure of situations, as well as for analyzing the stability of identified groups of situations. The methods of dynamic clustering are used to monitor the dynamics of changes in groups of situations, forecast their development depending on specific external influences represented by text descriptions.

- *situation analysis* • *neural network word processing* • *homogeneous semantic network*
- *fuzzy cognitive model* • *dynamic clustering*.



# Распознавание эмоций в персидской речи с помощью одномерной (1D) нейронной сети

*Сейедмилад Ранаей Сиадат,  
Московский физико-технический институт,  
Москва, Россия, ranaei.siadat.ss@phystech.edu*

Проблема распознавания и классификации эмоций в речи является одной из наиболее актуальных и значимых тем исследований; однако на сегодняшний день для большого количества языков практически не проводилось исследований с целью достижения требуемой точности [распознавания]. Выражение и распознавание эмоций на основе сигнала человеческой речи является одной из сложных проблем — помимо собственно языковой проблемы. В настоящей статье предлагается систематический и надежный подход к реализации системы распознавания эмоций для малоресурсных языков, таких как персидский. Насколько нам известно, это первая работа SER на материале персидского языка с использованием методов глубокого обучения. База данных эмоциональной речи Sharif ShEMO с пятью основными эмоциями (гнев, страх, счастье, печаль и удивление), а также нейтральным состоянием была определена как наиболее подходящий кандидат для оценки архитектуры одномерной сверточной нейронной сети (1DCNN). Данные сначала обрабатываются с использованием метода извлечения признаков коэффициентов косинусного преобразования Фурье для частот чистых тонов (MFCC), после чего MFCC поступают в качестве входных данных в нашу нейронную сеть. Результаты эксперимента показывают, что предлагаемый нами метод обеспечивает точность классификации [эмоций] около 74 % на наборе данных ShEMO.

• распознавание эмоций • обработка речи • эмоциональная речь • персидский язык • база данных персидской речи.

## I. ВВЕДЕНИЕ

За последние годы в отраслевых и академических кругах было проведено множество исследований по распознаванию человеческих эмоций, которые привлекли внимание многих ученых. Исследования в области систем ASR показали, что ключевыми элементами этих систем

являются лингвистические характеристики, таких как морфема, слово или предложение, что помогает лучше понять речь говорящего [1, 2]. Паралингвистические особенности, в том числе язык тела, жесты, мимика, тон и высота голоса, настроение и эмоции [3], а также лингвистическая информация являются двумя важными свойствами любого речевого сигнала со стороны говорящего; и человеческому мозгу необходимо и то и другое для лучшего понимания и эффективной коммуникации. Любой недостаток паралингвистических характеристик, таких как мимика и эмоции, негативно влияет на понимание и качество общения. Таким образом, для обеспечения эффективного и однозначно понятного общения, как у людей, важным фактором в развитии естественного взаимодействия человека с компьютером является потребность в паралингвистических характеристиках, таких как эмоции.

В настоящее время многие исследователи в области искусственного интеллекта пытаются изучать речь человека с целью анализа или выявления эмоционального состояния говорящего [4]. Выбор и извлечение этих признаков (характеристик) в ESR является сложной задачей [5, 6]. С целью классификации эмоций на основе их акустических характеристик (*энергии, формантных частот, основного тона частоты*) в речевых высказываниях [7, 8, 9] исследуются различные алгоритмы машинного обучения, такие как искусственные нейронные сети, машины опорных векторов, скрытые марковские модели (HMM) и гауссовы модели смешения. В извлечении этих признаков из речи и повышении эффективности и снижении вычислительной сложности при оценке человеческого поведения также сыграли важную роль подходы с применением глубокого обучения. С помощью методов глубокого обучения реализуются различные приложения в реальном времени для распознавания человеческих эмоций. Например, в случае взаимодействия человека и робота, если робот понимает эмоции человека по речи, для эффективного общения с говорящим могут быть адаптированы соответствующие действия; информация об эмоциональном состоянии учащихся может улучшить качество обучения и обеспечить более эффективные методы обучения; центры экстренного вызова могут определять эмоциональное состояние пользователя для обеспечения корректного реагирования. Существует множество моделей и методов в области глубокого обучения, таких как CNN [10], CNN-LSTM [11], которые используются для определения высокоуровневых признаков с целью достижения большей точности по сравнению с низкоуровневыми признаками при работе вручную. Эти методы также применяются к задаче ASR и показали современные характеристики эффективности. Однако эти исследования в основном сосредоточены на распознавании эмоций при анализе англоязычной речи и мандаринского наречия [китайского языка], и результаты не всегда применимы к анализу речи на других языках, таких как персидский.

В настоящей статье мы исследовали возможности одномерных сверточных нейронных сетей (1DCNN) в классификации речевых эмоций на персидском наборе данных, называемом базой данных эмоциональной речи Sharif (ShEMO) [12]. ShEMO — это полустественный набор данных, который включает образцы речи 87 носителей персидского языка для пяти наиболее часто используемых эмоций, а именно: *гнев, страх, счастье, печаль и удивление*, а также *естественное* состояние. Сначала из каждого аудиосигнала извлекаются характеристики коэффициентов косинусного преобразования Фурье для частот чистых тонов (MFCC), после чего вектор MFCC поступает в одномерную сверточную нейронную сеть (1D сверточную нейронную сеть) и используется SoftMax для классификации эмоций.

## II. СОПУТСТВУЮЩИЕ РАБОТЫ

На сегодняшний день учеными разработано множество методов и приемов выявления эмоционального состояния по цифровым звуковым сигналам в области SER. Как упоминалось в предыдущей главе, извлечение и выбор отличительных и характерных особенностей звукового сигнала для обнаружения эмоций — одна из самых сложных задач [13]. Недавние исследования в области SER показывают, что эту область науки можно разделить на две важные части.

1. Извлечение отличительных и высокоуровневых признаков из речи.
2. Выбор подходящего алгоритма для классификации этих признаков и распознавания эмоций по речи.

В последние десятилетия было проведено много исследований в области искусственного интеллекта и глубокого обучения, которые привели к решению сложных вычислительных задач для исследователей. Многие ученые в области SER разработали множество алгоритмов и методов глубокого обучения для повышения точности распознавания [эмоций] с использованием звуковых речевых сигналов (аудиосигналов). Как уже упоминалось, в рамках задач SER было проведено много исследований по извлечению ряда признаков из аудиосигнала и использованию сверточной нейронной сети (CNN) для классификации этих признаков и распознавания эмоций в речи. Например, Н. Суган [14] сравнил три различных кепстральных признака, таких как коэффициенты косинусного преобразования Фурье для частот чистых тонов (MFCC), коэффициенты косинусного преобразования Фурье с учетом человеческого фактора (HFCC) и коэффициенты косинусного преобразования Фурье гамматонной частоты (GFCC), и добился немного более высоких показателей на HFCC, чем на GFCC и MFCC, для двух разных наборов данных EMODB и SAVEE. Дейв и др. [15] оценили и сравнили различные типы извлечения признаков для речевых эмоций и показали наилучшие характеристики для признаков MFCC для SER, чем для других низкоуровневых признаков, таких как коэффициенты линейного предсказания (LPC), громкость и форманта. Цунь-Лонь [16] оценил извлечение пяти различных признаков, таких как LPC, MFCC, коэффициенты косинусного преобразования Фурье линейного прогнозирования (LPCC), логарифмические частотные коэффициенты мощности (LFPC), перцептивное линейное прогнозирование (PLP) на мандаринском наречии китайского языка [17], и использовал линейный дискриминантный анализ (LDA), метод k ближайших соседей (K-NN) и скрытые марковские модели (HMM) как классификаторы для классификации эмоций. Эндрю Хуань [18] представил методику, основанную на стандартной архитектуре CNN, аналогичную Alexnet [19], для глубокого обучения и обучил классификатор softmax распознавать эмоциональное состояние говорящего в наборе данных RAVDESS [20]. При этом сообщалось о максимальной точности в 85 % — наилучшей за 500 эпох с использованием расщепления данных испытаний 60/20/20.

Как упоминалось выше, из-за отсутствия данных и достаточных ресурсов на некоторых языках, таких как персидский, проведено не так много

исследований по извлечению данных и внедрению различных глубоких нейронных сетей на персидском наборе данных, подобном ShEMO. Омид Мохамад Незами [12] использовал расширенный женеvский минималистический набор акустических параметров (eGeMAPS) [21] и извлек эти признаки с помощью openSMILE [22] из набора данных ShEMO. Затем он нормализовал эти характеристики с использованием z-оценки. Он использовал три различных типа алгоритмов машинного обучения, таких как SVM с ядром радиальной базисной функции (RBF), (k-NN) и дерево решений (DT) с алгоритмом случайного леса для классификации. Кроме того, применялась 10-кратная перекрестная проверка с использованием байесовской оптимизации [23], применяемой для изменения параметров каждого классификатора. Он достиг точности 58,2 % для классификатора SVM, а также точности 47,6 и 47,4 % для K-NN и дерева решений соответственно.

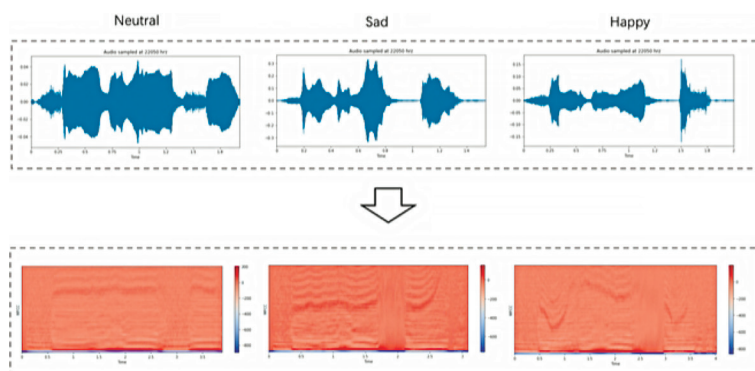


Рис. 1. Визуализированные MFCC трех различных эмоций (нейтрального состояния, грусти и счастья)

В настоящей работе для извлечения признаков из каждого сигнала были выбраны MFCC, которые можно рассматривать как представление аудиоданных в виде изображения по осям частоты и времени. Также оценивалась эффективность алгоритмов глубокого обучения, а не алгоритма машинного обучения, такого как SVM, K-NN и дерево решений.

### III. ЭКСПЕРИМЕНТАЛЬНАЯ МОДЕЛЬ

#### A. Сбор данных

Для проверки нашей методики и архитектуры использовалась база данных эмоциональной речи Sharif (ShEMO), представленная Омидом Мохамадом Незами [12]. База данных эмоциональной речи Sharif (ShEMO) — это общедоступная база данных персидской речи, которая содержит 3000 естественных высказываний в формате .wav (16 бит), записанных с частотой дискретизации 44,1 кГц в монозвук, а также извлеченные высказывания радиоспектаклей, которые транслируются онлайн, длительностью 3 часа и 25 минут. База данных ShEMO содержит образцы речи 87 носителей персидского языка (31 женщина, 56 мужчин) и включает аудиофайлы с пятью эмоциональными состояниями, такими как *гнев*, *страх*, *счастье*, *печаль* и *удивление*, а также *нейтральное состояние* вместе с орфографической и фонетической транскрипцией.



## В. Предварительная обработка данных

Предварительная обработка является одним из основных этапов подготовки входного сигнала для достижения [требуемой] точности и эффективности модели. Сначала мы использовали python 3.6 и библиотеку LibROSA (версия 0.7.2) [24] для загрузки аудиофайлов. Мы использовали исходную частоту дискретизации (22 Гц) и битовую глубину для обеспечения согласованной частоты дискретизации во всех базах данных, при этом звук предоставляется в стереоформате и нормализован в моноформате. После передискретизации всех высказываний мы очищаем аудиосигналы, чтобы удалить участки тишины из каждого речевого сигнала. Из каждого очищенного высказывания в наборе данных мы извлекаем признаки коэффициентов косинусного преобразования Фурье для частот чистых тонов, называемых MFCC (50 коэффициентов). В качестве оконной функции мы выбрали окно Ханна. Размер окна составляет 512 (23 миллисекунды при частоте дискретизации 22 050 Гц), а длина скачка — 256, т. е. перекрытие составляет 50 %, что обеспечивает хороший баланс между требуемыми характеристиками и вычислительной сложностью. Форма вывода MFCC для каждого высказывания представляет собой двумерный матричный формат с полосами MFCC по оси y и временем по оси x для отражения полос MFCC во времени. Мы визуализировали MFCC трех разных эмоций, таких как нейтральное состояние, грусть и счастье, которые показаны на рисунке 1.

## С. Обучение и тестовые выборки

Сначала мы разделили целые наборы данных shEMO для целей обучения и тестирования с 80 % и 20 % данных соответственно (данные обобщены в таблице 1). Затем мы вычислили признаки MFCC (50 коэффициентов) для каждого высказывания в обучающем наборе данных, а также в тестовой выборке данных. Наконец, метки обучающих и тестовых данных были закодированы как векторы прямого унитарного кода и переданы в нашу сеть. Следует отметить, что мы исключили высказывания с эмоцией страха в базе данных для обучения и тестирования, поскольку в базе данных их было немного (всего 38).

Таблица 1

**Число обучающих и тестовых выборок после расщепления 80/20 % по каждому эмоциональному состоянию**

Эмоция	ShEMO		
	Обучение (80 %)	Тест (20 %)	Всего
Счастье	165	36	201
Грусть	369	80	449
Гнев	829	230	127
Страх	32	6	38
Нейтральное	823	205	1028
Удивление	182	43	225

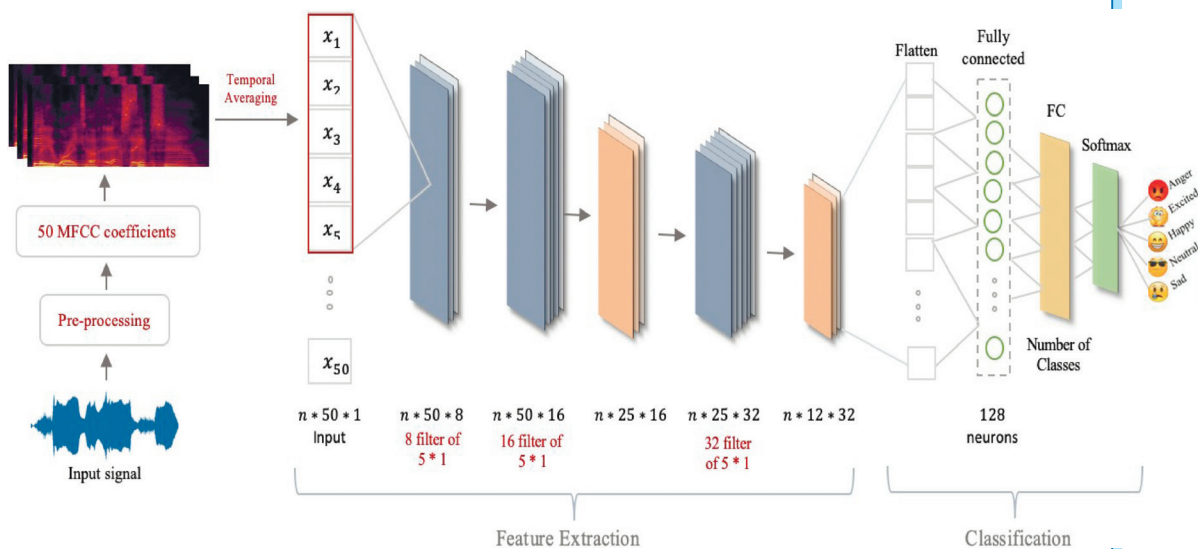


Рис. 2. Предлагаемая архитектура одномерной сверточной нейронной сети (1D CNN), используемая в текущем исследовании

#### D. Архитектура одномерной сверточной нейронной сети

Предлагаемая архитектура глубокой нейронной сети для SER показана на рисунке 2.

Архитектура глубокой нейронной сети, которая была реализована в нашем исследовании, представляла собой одномерную сверточную нейронную сеть (1D CNN) с тремя сверточными слоями и одним полносвязным слоем с 128 скрытыми нейронами. Как мы упоминали ранее, shEMo содержит аудиофайлы с пятью эмоциональными состояниями, такими как гнев, страх, счастье, печаль и удивление, а также нейтральным состоянием. Поэтому мы использовали softmax в качестве функции активации для конечных выходных узлов, чтобы оценить распределение вероятностей классов, каждый из которых принадлежит к одной из эмоций. Мы используем подвыборку с определением максимального значения после второго и третьего сверточного слоя. В качестве функции активации мы выбрали блоки линейной ректификации (ReLU) в сверточных и полносвязных слоях, чтобы снизить вероятность возникновения проблемы подстройки (переобучения) и ввести нелинейность в модель. Начальное количество ядер было установлено равным 8 и 16 для первого и второго сверточных слоев и 32 для третьего сверточного слоя соответственно. Кроме того, размер каждого сверточного фильтра составляет  $5 \cdot 1$  с шагом 1 и «такими же» интервалами заполнения для создания карт признаков. Размер ядра для подвыборки с определением максимального значения — 2 с шагом 2 и «такими же» интервалами заполнения. Мы использовали инициализацию Ксавье в полносвязном слое для настройки начальных значений веса [25]. В качестве функции потерь использовалась минимизация перекрестной энтропии, а в качестве блока оптимизации (оптимизатора) — Adam (адаптивная оценка моментов) с параметрами  $\text{lr}=0,0001$  [26]. Количество итераций обучения составляло 500. Кроме того, мы включили операцию отсева (исключения) после третьего сверточного слоя, чтобы предотвратить подстройку и улучшить характеристики.

Таблица 2

### Гиперпараметры и настройки предлагаемой сети

Параметр	Значение
Функция активации	Блоки линейной ректификации (ReLU)
Функция потерь	Минимизация перекрестной энтропии
Оптимизатор	Adam
Шаг (интенсивность) обучения	0.0001
Число эпох	500
инициализатор	Инициализация Ксавье

#### Е. Показатели оценки

После реализации алгоритма глубокого обучения и построения модели необходимо оценить характеристики (качество) модели перед тестированием модели на реальных данных, чтобы свести к минимуму ошибку. Показатели оценки помогают нам повысить качество модели глубокого обучения. Существует несколько видов параметров (метрик) для оценки моделей. Мы выбрали следующие матрицы оценки, такие как матрица неточностей, оценка (мера) f1, точность и отклик модели, чтобы повысить точность нашей модели. Сокращения TP и TN представляют собой истинно положительное и истинно отрицательное значение, аналогичным образом FP и FN представляют ложноположительное и ложноотрицательное значение.

**Скорость/правильность** классификации. Правильность — это отношение правильно спрогнозированного наблюдения к общему количеству наблюдений. Скорость или правильность классификации рассчитывается по следующему уравнению:

$$\text{Правильность} = \frac{(TP+TN)}{TP+TN+FP+FN} \quad (1)$$

**Точность** — это отношение положительных правильно спрогнозированных наблюдений к общему количеству спрогнозированных положительных наблюдений. Точность рассчитывается по следующему уравнению:

$$\text{Точность} = \frac{(TP)}{TP+FP} \quad (2)$$

**Отклик или чувствительность.** Это отношение правильно спрогнозированных положительных наблюдений к общему количеству положительных наблюдений. Отклик рассчитывается по следующему уравнению:

$$\text{Отклик} = \frac{(TP)}{TP+FN} \quad (3)$$

Оценка F1 представляет собой средневзвешенное значение точности и отклика и рассчитывается по следующему уравнению:

$$\text{Оценка } F1 = 2 \cdot \frac{\text{Точность} \cdot \text{Отклик}}{\text{Точность} + \text{Отклик}} \quad (4)$$

**Матрица неточностей** представляет собой таблицу степеней N\*N (N — количество классов), которая используется для обобщения результатов нашего прогноза с помощью числовых оценок в задаче классификации, истинные значения которых известны. В основном столбцы в матрице неточностей включают в себя фактические классы и классы прогнозирования строк, которые позволяют нам вычислять оценку f1 и показатели точности и отклика.

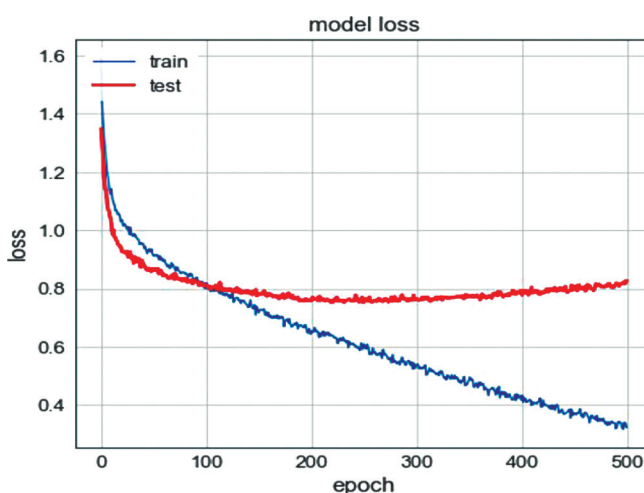


Рис. 3. Обучение vs ошибка тестирования на 500 эпох

#### IV. РЕЗУЛЬТАТ

Мы провели несколько экспериментов, чтобы повысить качество обучения и точность тестирования на базе данных ShEMO. Мы не смогли улучшить качество нашей модели путем изменения гиперпараметров, такие как размер ядер свертки и отсев. На рисунке 3 показан «потерянный» график между обучающимися и тестовыми выборками на 100 эпох. Согласно результатам, потери минимизируются после каждой эпохи. Качество обучения и тестирования начинают совпадать примерно через 100 эпох. Общая точность тестирования составляет 71,9 %. На рисунке 4 показан «потерянный» график между обучающимися и тестовыми выборками на 500 эпох. Примерно после 120-й эпохи графики обучения и тестирования пересеклись друг с другом, что означает переобучение. Из-за этого график тестирования после точки пересечения не улучшится, даже если улучшится качество обучения. Качество обучения будет продолжать улучшаться, даже если модель обучается на большем числе эпох. Кроме того, применение регуляризации Ксавье в полносвязном слое помогает сгладить графики точности и тестирования. Общая точность тестирования составляет 73,7 %. На рисунках 5 и 6 показаны нормализованные матрицы неточностей для модели, выполненные для 100 и 500 эпох соответственно. Что касается отдельных эмоций, то нейтральное состояние и гнев достигли

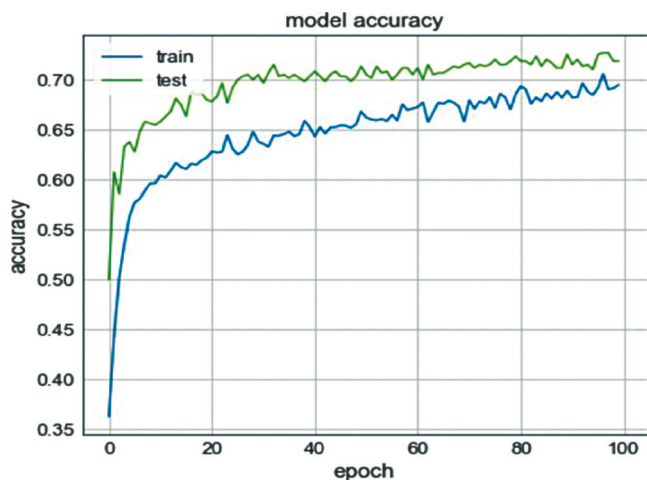


Рис. 4. Обучение vs ошибка тестирования на 100 эпох

наивысшей точности со значениями 0,83 и 0,84 соответственно при прогоне на 500 эпох. В основном это связано с тем, что эти две эмоции имеют наибольшее количество образцов в наборе данных. Напротив, эмоция с наименьшим количеством образцов (это счастье) достигла самой низкой индивидуальной точности со значением 0,22. Согласно [12], эмоцию счастье чаще всего путают с гневом. Однако результат здесь показывает, что счастье в основном классифицируется как нейтральное состояние и реже как гнев. Это означает, что примененная здесь модель может лучше различать, что счастье и гнев являются эмоциями разного спектра, а не просто двумя похожими эмоциями сильного возбуждения. Это похоже на правду, поскольку нейтральное состояние ближе к счастью в эмоциональном спектре по сравнению с гневом. Несмотря на то что общая точность тестирования одинакова, т. е. составляет 71,9 и 73,7 %, матрица неточностей для 500 эпох показывает лучший результат.

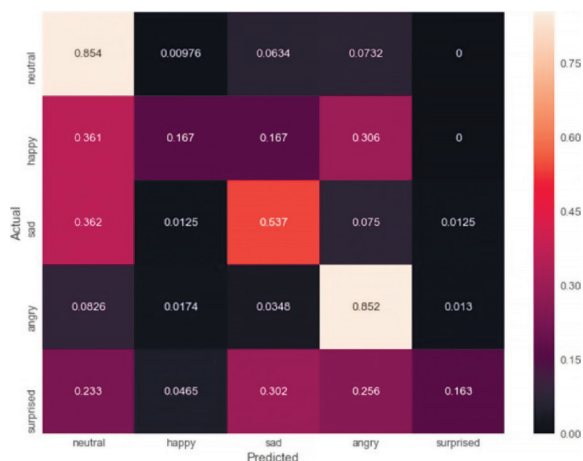


Рис. 5. Точность тестирования нормализованной матрицы неточностей для базы данных ShEMO для 500 эпох

Для сравнения, значение эмоции удивления увеличилось с 0,16 до 0,40, а значение эмоции счастья увеличилось с 0,16 до 0,22. Несмотря на то что использование 500 эпох дает лучшую точность и матрицу неточностей, это может быть верно только для этого набора данных, а результат может быть хуже для неотображаемых данных из-за переобучения, которое произошло примерно после 120-й эпохи. Это важно отметить, особенно для практического применения. Еще один важный аспект, который следует отметить, это то, что эмоция страха полностью исключена из набора данных, поскольку он включает всего 38 ее образцов.

В таблице ниже показаны соответствующие характеристики для 594 тестовых образцов по точности, отклику и оценке f1 в независимом от пола испытуемых (гендерно-независимом) режиме. Показано, что общая оценка f1 составляет 74 % на 500 эпох с использованием расщепления 80/20 данных тестирования для пяти эмоций (нейтральное состояние, счастье, грусть, гнев и удивление). Согласно нашей точности, предложенная модель в гендерно-независимом режиме достигла результата с точностью на 15,5 % лучше, чем классификатор SVM Омида Мохамеда Незами [12] с точностью 58,2 %. Как показано, первое наблюдение заключается в том, что у гнева лучшая точность и оценка f1. С другой стороны, нейтральное состояние дало лучший отклик. Оба признака имеют наибольшее количество образцов, как указано соответственно.

Таблица 3

### Результаты эксперимента на базе ShEMO для 1D CNN

	Точность	Отклик	Оценка F1	опорное значение
Счастье	165	36	201	36
Грусть	369	80	449	80
Гнев	829	230	127	230
Нейтральное состояние	823	205	1028	205
Удивление	182	43	225	43
Итого	0.74	0.74	0.74	594

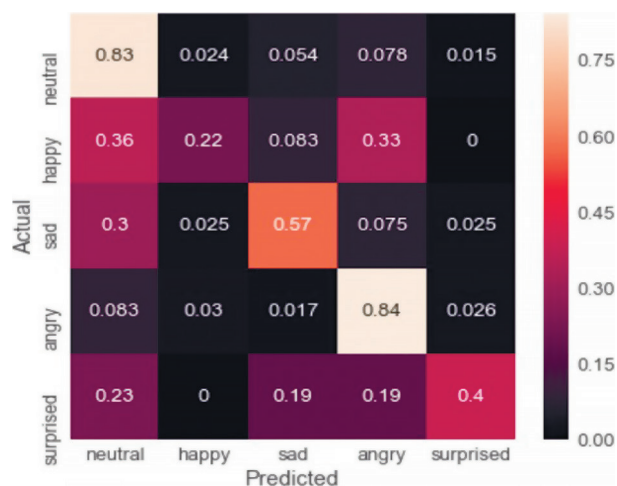


Рис. 6. Точность тестирования нормализованной матрицы неточностей для базы данных ShEMO для 100 эпох

## V. ИТОГИ И ВЫВОДЫ

Задача распознавания речевых эмоций на персидском языке решается 1DCNN на базе данных ShEMO. ShEMO — это набор речевых высказываний на персидском языке, состоящий из 3000 образцов, разделенных на шесть различных эмоций. Сначала мы использовали исходную частоту дискретизации (22 Гц) и битовую глубину для обеспечения постоянной частоты дискретизации во всех базах данных, при этом аудиоданные предоставляются в стереоформате и нормализуются в моноформате. Затем из речевых высказываний извлекаются признаки с использованием MFCC с 50 коэффициентами. После этого признаки MFCC сжимаются путем усреднения признаков во времени, поэтому аддитивный шум речевого фрагмента снижается. Теперь, когда входные данные обработаны надлежащим образом, они распространяются через предложенную архитектуру модели, состоящую из трех сверточных слоев одномерной сети, одного полносвязного слоя со 128 нейронами и выходного слоя SoftMax, количество нейронов которого зависит от выходных меток. Сеть прошла обучение в течение 500 эпох для данных ShEMO. В процессе обучения имело место переобучение (подстройка) с 500 эпохами, но необходимо продолжить работу, чтобы повысить качество для набора данных ShEMO. С точки зрения только точности по сравнению с другими известными работами, предложенная архитектура модели превзошла классификатор SVM Омида Мохамеда Незами [12] на 15,5 % в случае набора данных ShEMO. Это показывает, что предложенное решение является лучшим для ShEMO и может рассматриваться как новое современное решение, по крайней мере, для набора данных ShEMO, поскольку результат Омида Мохамеда Незами [12] был предыдущим лучшим результатом. Это демонстрирует, что предложенная модель успешно преодолела сложность решения задачи распознавания эмоций в речи с использованием примененного метода извлечения признаков.

## VI. ДАЛЬНЕЙШИЕ ИССЛЕДОВАНИЯ

Даже при таких отличных результатах возможно дальнейшее повышение качества предложенной архитектуры. Одним из потенциальных улучшений, которое можно реализовать, является использование различных методов извлечения признаков, таких как коэффициенты линейного предсказания (LPC) и кепстральные коэффициенты линейного предсказания (LPCC). Еще одно усовершенствование, которое может дать лучший результат, заключается в реализации другой архитектуры нейронной сети вместо 1DCNN, такой как двунаправленная LSTM и другие гибридные сети. Кроме того, также интересно дополнительно изучить качество предлагаемой модели на объединенных наборах данных в рамках одного языка.

### Литература

1. Juang B.-H., Levinson S.E., Rabiner L.R., and Sondhi M.M. Hidden markov model speech recognition arrangement. Nov. 8 1988, uS Patent 4,783,804.

2. *Levinson S.E.* Continuously variable duration hidden markov models for automatic speech recognition. *Computer Speech & Language*, vol. 1, no. 1, pp. 29–45, 1986.
3. *Roach P., Stibbard R., Osborne J., Arnfield S. and Setter J.* Transcription of prosodic and paralinguistic features of emotional speech. *Journal of the International Phonetic Association*, vol. 28, no. 1–2, pp. 83–94, 1998.
4. *Grewe L. and Hu C.* Ulearn: understanding and reacting to student frustration using deep learning, mobile vision and nlp. In *Signal Processing, Sensor/Information Fusion, and Target Recognition XXVIII*, vol. 11018. International Society for Optics and Photonics, 2019, p. 110180W.
5. *Chaudhary A., Sharma A.K., Dalal J and Choukiker L.* Speech emotion recognition. *Journal of Emerging Technologies and Innovative Research*, vol. 2, no. 4, pp. 1169–1171, 2015.
6. *Wei B., Hu W., Yang M. and Chou C.T.* From real to complex: Enhancing radio-based activity recognition using complex-valued csi,” *ACM Transactions on Sensor Networks (TOSN)*, vol. 15, no. 3, pp. 1–32, 2019.
7. *Jain M., Narayan S., Balaji P., Bhowmick A., Muthu R.K. et al.* Speech emotion recognition using support vector machine. *ArXiv preprint arXiv:2002.07590*, 2020.
8. *El Ayadi M.M., Kamel M.S. and Karray F.* Speech emotion recognition using gaussian mixture vector autoregressive models. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*, vol. 4. IEEE, 2007, pp. IV–957.
9. *Han K., Yu D. and Tashev I.* Speech emotion recognition using deep neural network and extreme learning machine. In *Fifteenth annual conference of the international speech communication association*, 2014.
10. *Krizhevsky A., Sutskever I. and Hinton G.E.* Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, 2012, pp. 1097–1105.
11. *Hochreiter S. and Schmidhuber J.* Long short-term memory. *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
12. *Nezami O.M., Lou P.J. and Karami M.* Shemo: a large-scale validated database for persian speech emotion detection. *Language Resources and Evaluation*, vol. 53, no. 1, pp. 1–16, 2019.
13. *Salwa R.F.G. Al-agma A. , Hilal H. Saleh.* Analyze features extraction for audio signal with six emotions expressions. *International Journal of Engineering and Advanced Technology (IJEAT)*, 2015.
14. *Sugan N., Srinivas N.S., Kar N., Kumar L., Nath M.K. and Kanhe A.* Performance comparison of different cepstral features for speech emotion recognition. In *2018 International CET Conference on Control, Communication, and Computing (IC4)*, 2018.
15. *Dave N.* Feature extraction methods lpc, plp and mfcc in speech recognition. *International journal for advance research in engineering and technology*, vol. 1, no. 6, pp. 1–4, 2013.
16. *Pao T.-L., Chen Y.-T. and Yeh J.-H.* Emotion recognition from mandarin speech signals. In *2004 International Symposium on Chinese Spoken Language Processing. IEEE*, 2004, pp. 301–304.
17. *H. Bu, J. Du, X. Na, B. Wu, and H. Zheng.* Aishell-1: An open-source mandarin speech corpus and a speech recognition baseline. In *2017 20th Conference of the Oriental Chapter of the International Coordinating Committee on Speech Databases and Speech I/O Systems and Assessment (O-COCOSDA)*. IEEE, 2017, pp. 1–5.
18. *Huang A. and Bao P.* Human vocal sentiment analysis. *ArXiv preprint arXiv:1905.08632*, 2019.
19. *Alom M.Z., Taha T.M., Yakopcic C., Westberg S., Sidike P., Nasrin M.S., Van Esesn B.C., Awwal A.A.S. and Asari V.K.* The history began from alexnet: A comprehensive survey on deep learning approaches. *ArXiv preprint arXiv:1803.01164*, 2018.
20. *Livingstone S.R. and Russo F.A.* The ryerson audio-visual database of emotional speech and song (ravdess): A dynamic, multimodal set of facial and vocal expressions in north american english. *PLoS one*, vol. 13, no. 5, 2018.
21. *Eyben F., Scherer K.R., Schuller B.W., Sundberg J., André E., Busso C., Devillers L.Y., Epps J., Laukka P., Narayanan S.S. et al.* The geneva minimalistic acoustic parameter set (gemaps) for voice research and affective computing. *IEEE transactions on affective computing*, 2015.



22. *Eyben F., Wöllmer M. and Schuller B.* Opensmile: the munich versatile and fast open-source audio feature extractor. In Proceedings of the 18th ACM international conference on Multimedia, 2010, pp. 1459–1462.
23. *Snoek J., Larochelle H. and Adams R.P.* Practical bayesian optimization of machine learning algorithms. In Advances in neural information processing systems, 2012, pp. 2951–2959.
24. Librosa library. [Online]. Available: <https://librosa.github.io/librosa/>
25. *Glorot X. and Bengio Y.* Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the thirteenth international conference on artificial intelligence and statistics, 2010, pp. 249–256.
26. *Kingma D.P. and Ba J.* Adam: A method for stochastic optimization. ArXiv preprint arXiv:1412.6980, 2014.

## EMOTION RECOGNITION IN PERSIAN SPEECH USING A ONE-DIMENSIONAL (1D) NEURAL NETWORK

*Seyedmilad Ranaei Siadat,  
Moscow Institute of Physics and Technology,  
Moscow, Russia, ranaei.siadat.ss@phystech.edu*

The problem of recognition and classification of emotions in speech is one of the most relevant and significant research topics; however, to date, for a large number of languages, almost no research has been carried out in order to achieve the required accuracy [of recognition]. The expression and recognition of emotions based on the signal of human speech is one of the difficult problems — in addition to the language problem itself. This paper proposes a systematic and robust approach to implementing an emotion recognition system for low-resource languages such as Persian. To the best of our knowledge, this is SER's first work in Persian using deep learning methods. The Sharif ShEMO Emotional Speech Database, with five primary emotions (anger, fear, happiness, sadness, and surprise) as well as a neutral state, was identified as the most suitable candidate for evaluating a 1D Convolutional Neural Network (1DCNN) architecture. The data is first processed using the pure tone cosine Fourier transform coefficient (MFCC) feature extraction method, after which the MFCCs are fed as input to our neural network. The results of the experiment show that our proposed method provides a classification accuracy [of emotions] of about 74 % on the ShEMO dataset.

• *emotion recognition* • *speech processing* • *emotional speech* • *Persian language* • *Persian speech database*.

# Когнитивные исследования в режиме реального времени: нейросетевой подход<sup>1</sup>

*Пильгун М.А., Российский государственный социальный университет, Институт языкознания РАН,  
Москва, Россия, pilgunm@yandex.ru*

В статье представлены результаты когнитивного исследования, посвященного анализу взаимоотношений жителей и властей города с использованием нейросетевого подхода. Исследование проводилось на примере градостроительного конфликта в г. Москве в режиме реального времени. Материалом послужили данные из социальных сетей, микроблогов, блогов, мессенджеров, форумов, обзоров, видеохостингов, тематических порталов, интернет-СМИ, печатных СМИ и телевидения, связанные со строительством транспортно-пересадочного узла «Суворовская» в центре г. Москвы. Для анализа контента социальных медиа использовался мультимодальный подход, нейросетевой анализ речевых данных, сентимент-анализ, анализ лексических ассоциаций и контент-анализ. Динамика коммуникационных процессов в информационном пространстве вокруг реализации проекта характеризуется эскалацией конфликта между жителями, строителями и властями города. Анализ данных позволяет говорить о дальнейшем развитии конфликта по мере реализации проекта.

• когнитивные исследования • восприятие речи • данные социальных медиа • нейросетевые технологии.

## 1. ВВЕДЕНИЕ

Одна из актуальных проблем современной научной парадигмы — интерпретация Больших данных в реальном времени. Когнитивные исследования представляются наиболее перспективным для решения подобных задач, в частности адаптация психолингвистических методик.

Так, анализ восприятия речи в многоуровневой системе обработки представлен в достаточном количестве работ [1; 2; 3; 4; 5]. Развитие «глубокого обучения» привело к повышению эффективности компьютерных моделей восприятия, что в определённой степени пошатнуло убеждение в том, что восприятие является уникальной способностью биологических систем, внутренние механизмы которой еще недостаточно исследованы, а также что восприятие не может быть представлено у искусственных существ на уровне человека [6].

<sup>1</sup> Работа выполнена в рамках проекта РФФИ «Структура и содержание представлений о человеке, государстве, власти в языковом сознании русскоязычных (Республика Крым)» (19-012-00295 А).

Современные генеративные модели могут создавать текстовое описание и новые изображения, генерировать изображения из описаний естественного языка [7, 8]. Поведенческие и нейровизуализированные данные позволяют проанализировать восприятие и воспроизведение речи у здоровых и клинических групп населения, в разных возрастных группах [9]. Комбинирование методов нейробиологии и качества эксперимента показывает, что речевые стимулы разной длины при различных нарушениях степени устойчивости сопровождаются физиологическими реакциями, связанными с вариативностью, например с потенциальной возможностью положительного пика [10].

Между тем методики интерпретации данных социальных медиа, восприятия речи пользователей для определения общественного мнения еще нуждаются в дальнейшем совершенствовании. Кроме того, пандемия коронавируса радикально и изменила жизнь общества, и заставила перестроить методологию исследований, в частности анализа обработки естественного языка и производства речи в виртуальной среде. Необходимо изучить, как изменилось восприятие людей, понимание и реагирование на среду, взаимодействие друг с другом и т. д. [11].

Цель данной статьи — представить алгоритм исследования общественного мнения с помощью нейросетевого анализа восприятия речи акторов социальных медиа на примере градостроительного конфликта в г. Москве (РФ).

### 1.1 . Данные

Материалом для исследования послужили данные из социальных сетей, микроблогов, блогов, мессенджеров, форумов, обзоров, сервисов видеохостинга, тематических порталов, интернет-СМИ, печатных СМИ и телевидения, связанных со строительством транспортно-пересадочного узла (ТПУ) «Суворовская» в г. Москве (Россия).

В конце сентября 2019 года жителям Центрального округа г. Москвы стало известно о проекте строительства ТПУ «Суворовская», данная информация наложилась на развивающийся конфликт, связанный с реконструкцией спортивного комплекса «Олимпийский».

Количественные характеристики данных представлены в таблице 1.

Таблица 1

#### Количественные характеристики данных

Параметры	Количество
Количество сообщений	374
Максимум сообщений в сутки	54
Количество авторов	264
Активность (постов на автора)	1,42
Количество источников	44
Количество токенов	398 292

## 1.2. Метод

Для анализа контента социальных медиа использовался мультимодальный подход, а также нейросетевой анализ речевых данных, анализ тональности, анализ словесных ассоциаций [12], контент-анализ [13; 14]

Для интерпретации данных применялась нейросетевая технология TextAnalyst 2.3, построенная с помощью нейроподобных элементов с временным суммированием сигналов (кортикоморфная ассоциативная память), что позволяет выделить эксплицитные знания, ключевые темы, которые вызвали наибольший интерес акторов, изучить тематическую структуру и провести саммаризацию базы данных. Кроме того, нейросетевое представление текста позволяет сформировать и интерпретировать семантическую сеть в виде набора взаимосвязанных понятий. С помощью семантической сети были выделены наиболее важные для акторов имплицитные темы, проранжированы семантические акценты. Проведение ассоциативного поиска, анализ ассоциативных сетей релевантных стимулов позволили проанализировать восприятие речи акторов [12].

## 2. РЕЗУЛЬТАТЫ И ДИСКУССИЯ

### 2.1. Общая характеристика контента

По данным анализируемого периода первый всплеск информационной активности, связанный с проектом ТПУ «Суворовская», относится к 30.01.2019, был вызван твитом мэра С. Собянина о возобновлении строительства станции. Сообщение активно растиражировали традиционные и социальные медиа. Проект ТПУ «Суворовская» рассматривался в рамках реализации строительства БКЛ.

Интерес пользователей к проекту, рост динамики просмотров сообщений, определяющий интерес к теме широкого круга общественности, фиксируется только 07.08.2019 и может рассматриваться как реакция на новый всплеск информационной активности, вызванный ростом сообщений и активизацией авторов с 31.07.2019. Самый высокий пик активности авторов, роста сообщений представлен 04.02.2020, пик роста просмотров с 05.02.2020 (рис. 1, 2, 3).

Геолокация пользователей показывает достаточно широкий круг стран: Россия, Великобритания, Армения, Венгрия, Латвия, Украина, Беларусь, ОАЭ, США, Китай, Южная Корея, Япония, Монголия, Германия, Словакия, Франция, Швейцария. Закономерно, что большинство акторов находилось в Москве. Также заинтересованность в ходе реализации проекта проявили жители Санкт-Петербурга, Солнечногорска, Ярославля, Калуги, Нижнего Новгорода, Красногорска, Самары, Электростали, Екатеринбурга, Серпухова, Нефтекамска, Салавата, Уфы, Омска, Иркутска, Севастополя.

Наиболее активными акторами, генерирующими контент, посвященный ТПУ «Суворовская», оказались медийные ресурсы (рис. 4).

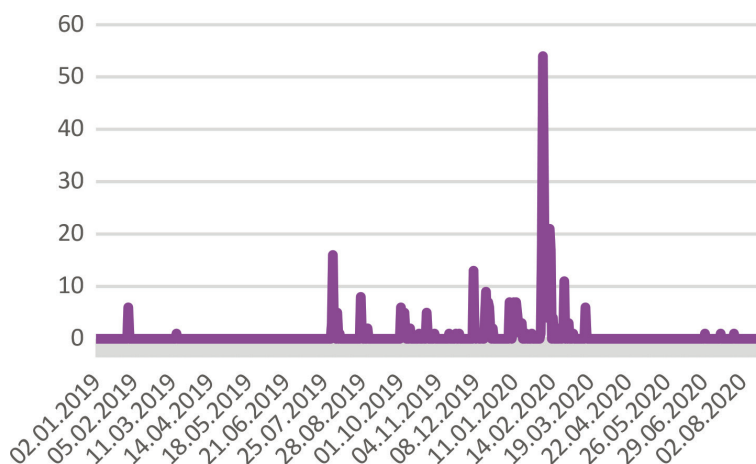


Рис. 1. Динамика сообщений

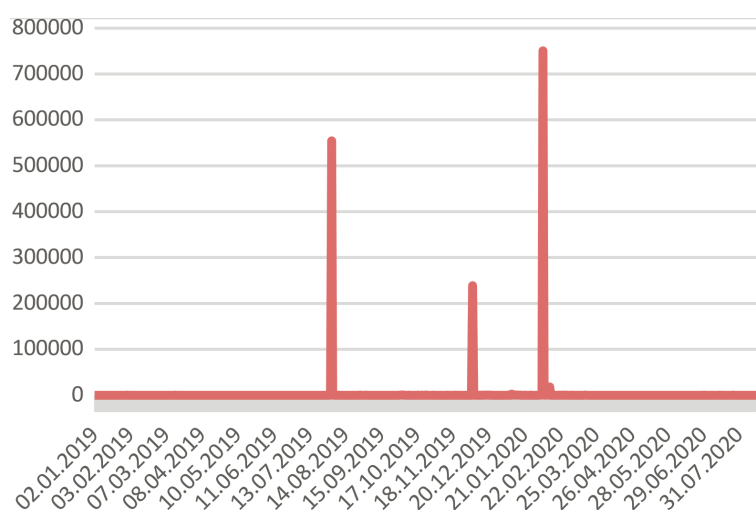


Рис. 2. Динамика просмотров

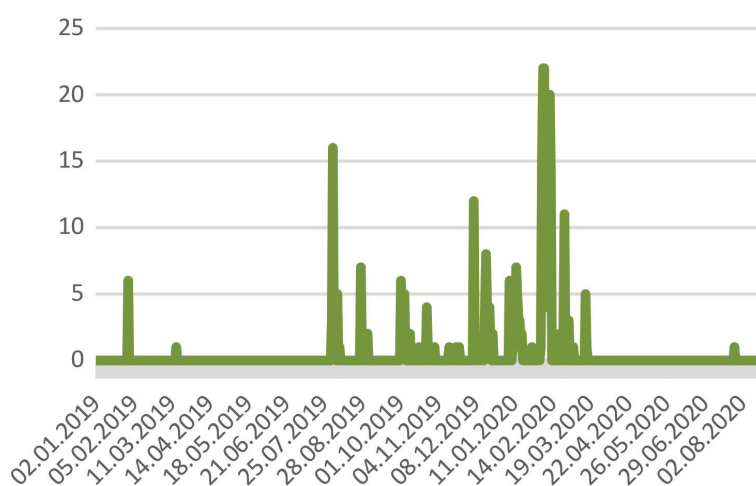


Рис. 3. Динамика активности авторов

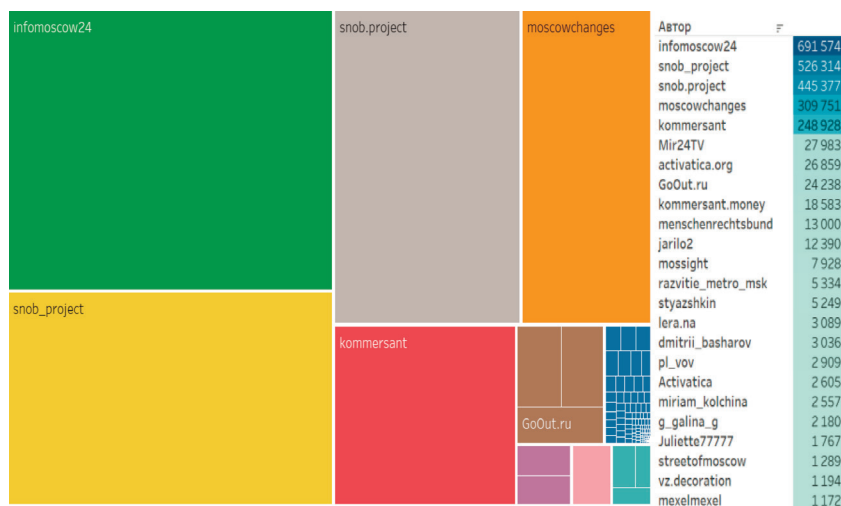


Рис. 4. Рейтинг авторов

Жители Мещанского и Тверского районов создали петицию против строительства высоток и ТПУ «Суворовская» на портале change.org, адресованную властям города, которую подписали 2940 пользователей (на 02.06.2021). Петиция активно распространяется через различные цифровые платформы.

Максимальное количество сообщений, раскрывающих различные аспекты реализации проекта ТПУ «Суворовская», представлено в микроблогах. Некоторое количество — в социальных сетях, незначительное количество — в интернет-СМИ, на тематических порталах, в мессенджерах, блогах, на форумах (рис. 5). Лидерами среди цифровых источников, которые предпочитали акторы, стали «Твиттер», «ВКонтакте» и «Инстаграм» (рис. 6).

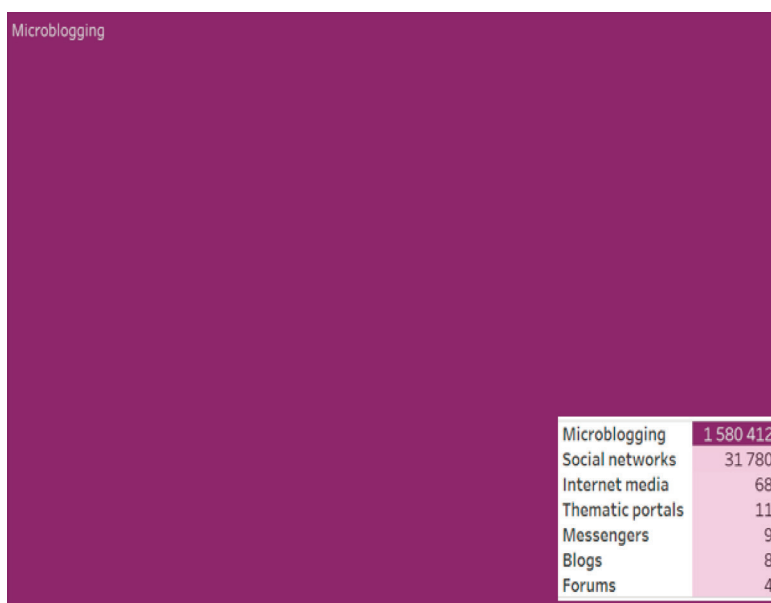


Рис. 5. Типы источников, которые использовали акторы



Рис. 6. Источники, которые использовали акторы

Между тем, непосредственная реакция пользователей (лайки, комментарии, просмотры) представлена в социальных сетях, которые могут служить основным источником при анализе восприятия пользователей (рис. 7).

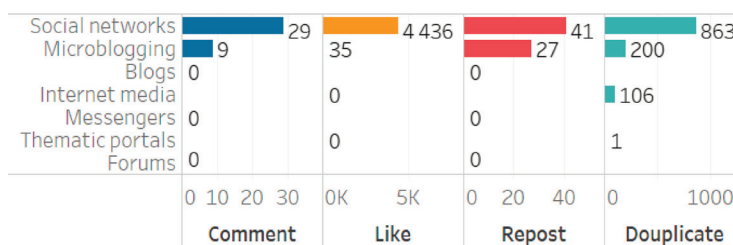


Рис. 7. Типы источников с цифровыми следами акторов

Наибольшей популярностью у авторов, освещающих реализацию проекта, пользуются «Твиттер» и «Фейсбук». Обсуждение результатов строительства происходит также в «ВКонтакте» и «Инстаграм». Рейтинг авторов с учетом использованных разных платформ показан на рис. 8.

Самый большой охват аудитории (724 673) получил официальный информационный текст с заголовком «Началось проектирование станции "Суворовская" Кольцевой линии метро» (Москва 24 (infomoscw24)).

Максимальное количество лайков (629) также получила аналогичная официальная информация:

*В столице стартовали работы по проектированию станции «Суворовская» Кольцевой линии, рассказал заместитель мэра Москвы по вопросам градостроительной политики и строительства Андрей Бочкарев.*

Между тем рейтинг текстов, получивших максимальное количество комментариев, возглавил текст пользователя с резко негативной оценкой проекта:

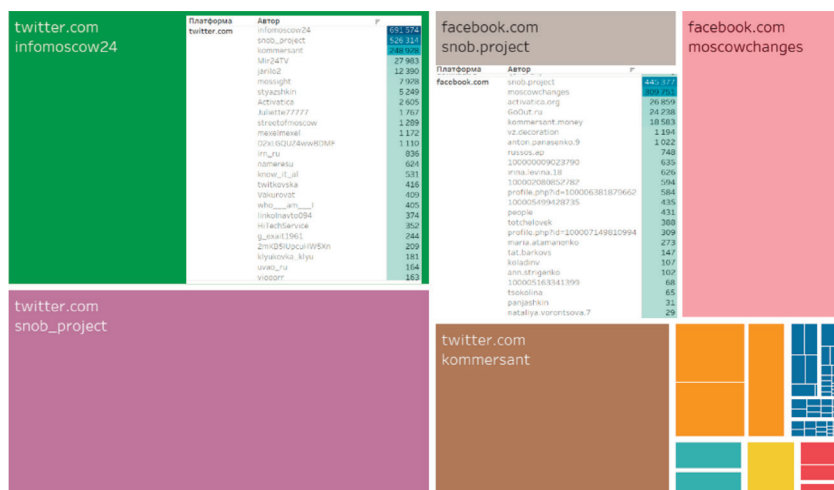


Рис. 8. Рейтинг авторов с учетом использованных разных платформ

Друзья, вечер перестаёт быть томным! Нас выселяют из нашего района! Нельзя оставаться равнодушными, чтобы потом не говорить, как же всё застроили.... более одного миллиона квадратных метров и парковка на 5000 машин около чудесной мечети, которая будет после застройки маленьким домиком! Москва-резиновая? Знаю, что наши голоса почти ничего не значат, но сидеть тихо тоже не вариант! Попробовать стоит! #сколимпийский #тпсуворовская #мещанскийрайон #застройкамосквы #моямосква #мой-район

Максимальное количество репостов также получил пост с резкой критикой строительства:

И такое в Москве возможно! Так называемая реконструкция спорткомплекса «Олимпийский» в Москве, а фактически новое строительство, проводится без разрешительных документов (...) ТПУ «Суворовская», стартовая площадка по программе реновации на улице Дурова, владение 3, и строительство высоток у спорткомплекса «Олимпийский». <bg>Всё это строительные работы рядом. То есть застройщики берут целый микрорайон и полностью, да ещё и без разрешений, меняют его облик. И так район за районом меняют Москву на неизвестный город.</bg>Проведение проверки по спорткомплексу «Олимпийский» уже согласовано Прокуратурой, добиваюсь проверок и по другим адресам в этом районе.<bg>#Москва#Олимпийский#Петербург#Мосгордума.

## 2.2. Тематическая структура и семантическая сеть контента

Анализ тематической структуры сводного датасета и результатов показал, что наибольший вес связи имеют понятия: *высотка* (84), *Суворовская Кольцевая линия* (86), *станции метрополитена* (80), *проектировать* (81), *строительство* (82), *район* (62), *Олимпийский* (59), *город* (57), *мэр* (55), *Проспект* (45). Контексты показали что, при обсуждении реализации проекта ТПУ «Суворовская» наиболее близкими проблемами пользователи считали вопросы строительства высоток, состояния транспортной системы, реконструкции комплекса «Олимпийский», развития района и городской среды, а также деятельность мэра.



Семантическая сеть сводного датасета демонстрирует семантические акценты, важные для жителей вопросы, где на первое место выходит недовольство проектом, протест, который был реализован в петиции на портале change.org «Жители Мещанского и Тверского районов против высоток и строительства ТПУ «Суворовская». Вес связи номинации «петиции» в семантической сети — 99.

В семантическом ядре весом связи 99 получают номинации, обозначающие проблемы, которые наиболее сильно беспокоят жителей: *строительство высоток, нарушение градостроительной политики, ограничение движения, функционирование метрополитена, реконструкция комплекса «Олимпийский»*.

В ходе обсуждения проблем строительства ТПУ «Суворовская», жители также активно критикуют деятельность мэра С. Собянина [99] и заместитель мэра А. Бочкарева [99].

## 2.3. Ключевые темы пользовательского контента

### 2.3.1. Негативный кластер

Негативный кластер составляют претензии жителей к властям города и строителям в связи с реализацией проекта.

- По мнению жителей, строительство ТПУ «Суворовская» необходимо властям города и строителям только для того, чтобы построить высотные дома. Проблема строительства ТПУ оказывается тесно связанной в сознании жителей с проблемами стартовой площадки по программе реновации города Москвы (улица Дурова, владение 3). Именно это вызывает наибольшие опасения жителей. Они обвиняют московские власти в нежелании устранить проблемы, уладить конфликт.
- Конфликт между жителями и городскими властями по поводу проекта ТПУ «Суворовская» накладывается на конфликт, связанный с реконструкцией спортивного комплекса «Олимпийский» (Олимпийский проспект, дом 16, стр. 1, 2, 3, 4). Строительство ТПУ и снос спортивного комплекса «Олимпийский» воспринимается как единая проблема.
- Повышение высотности, увеличению плотности застройки, следовательно, резкое ухудшение уровня жизни жителей района.
- Проект наносит ущерб природным парковым зонам, ведет к ухудшению экологической ситуации.
- Не учтены инженерные мощности, что приведет к необходимости перекладывать существующих коммуникаций.
- По мнению жителей, проектные показатели ТПУ имеют существенные недостатки, в частности не учтена пропускная транспортная ситуация прилегающих районов, таким образом, во время строительства возникнут серьезные ограничения движения, а после завершения — приведёт к транспортному коллапсу района.
- В результате реализации проекта Мещанский район полностью утратит исторический облик.

- Недостаточная проработка технических аспектов строительства.
- Нарушения при оформлении правовой и строительной документации.
- Манипуляции и обман на публичных слушаниях.
- Планируемая парковка на 5000 машино-мест значительно ухудшит транспортную и экологическую ситуацию в районе.

### 2.3.2. Позитивный кластер

Положительная тональность характеризует преимущественно официальные материалы, описывающие ход строительства ветки, темпы строительства, передовые методы, которые применяют строители. Как положительный момент ангажированными акторами подаётся информация о возобновлении старого проекта по развитию транспортной системы города, об улучшении инфраструктуры района, а также тот факт, что при строительстве ТПУ «Суворовская» будут решены сложные инженерные задачи.

### 2.4. Формирование рейтинга социальной напряженности вокруг реализации проекта ТПУ «Суворовская»

На протяжении анализируемого периода можно отметить значительный рост сообщений, посвящённых реализации проекта ТПУ «Суворовская» (рис. 9).

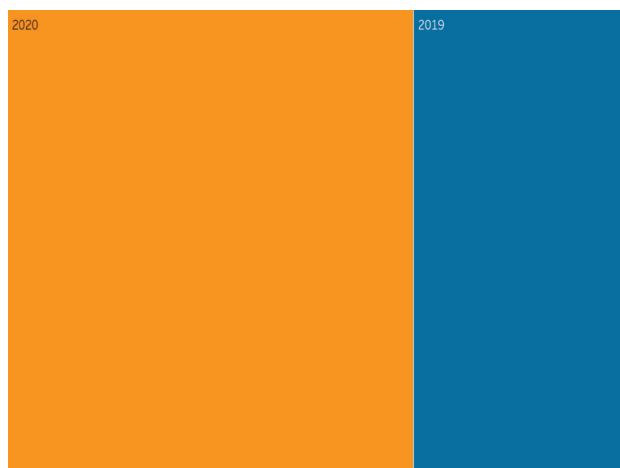


Рис. 9. Рост сообщений, посвящённых реализации проекта ТПУ «Суворовская»

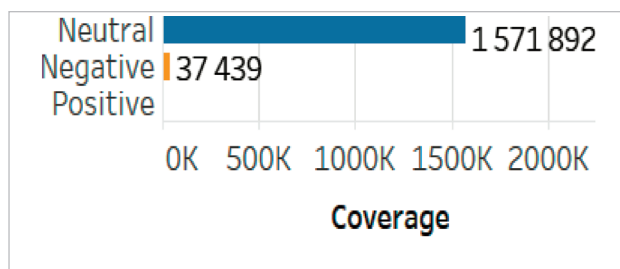


Рис. 10. Тональность охвата

В сводной базе данных по охвату аудитории преобладает нейтральный кластер (рис. 10). Однако следует учитывать, что значительное количество нейтральных и позитивных упоминаний определяется увеличением числа официальных материалов. Между тем динамика тональности сообщений, посвященных строительству ТПУ «Суворовская», за 2019–2020 гг. демонстрирует рост негативных сообщений и уменьшение позитивных упоминаний (рис.11). Рейтинг цифровых следов по тональности также показывает преобладание негативных комментариев, дублей и репостов (рис.12). Подобная ситуация характеризует рост негативного отношения пользователей к проблеме.

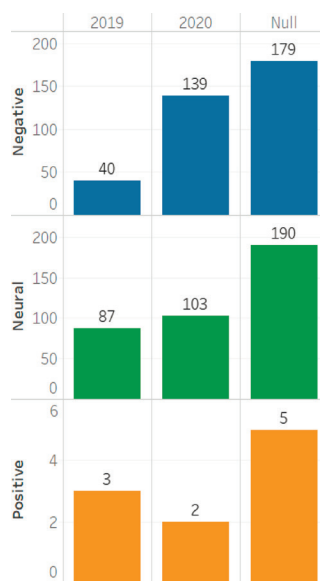


Рис. 11. Динамика упоминаний с разной тональностью

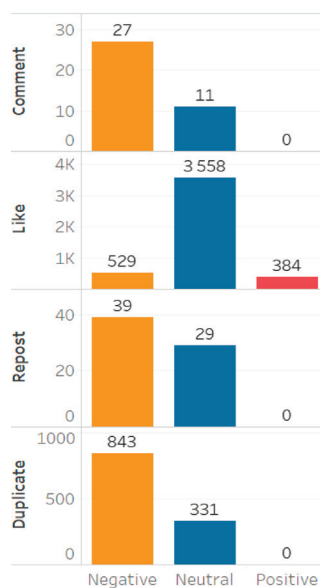


Рис.12. Динамика цифровых следов акторов с разной тональностью

## 2.5. Выявление наличия/отсутствия социального стресса в районе строительства ТПУ «Суворовская»

Выявление наличия/отсутствия социального стресса в районе строительства ТПУ «Суворовская» по цифровым данным проводилось с использованием нейросетевых технологий по методике, описанной в [12].

Специфика данного объекта заключается в том, что это локальный конфликт, который затрагивает интересы жителей Центрального округа. Поскольку конфликт территориально ограничен, активисты имеют возможность часто собираться и лично обсуждать проблемы.

Сочувствия большого числа пользователей данной ситуации не вызывает, поскольку проблемы благополучных жителей Мещанского и Тверского районов не затрагивают жителей отдаленных спальных районов или Новой Москвы.

Ядро минимального древовидного подграфа исходной ассоциативной сети, построенный на вершине, соответствующей стимулу «ТПУ «Суворовская», по данным негативного кластера, имеет вес 10/551. В центре оказывается петиция на портале change.org, в которой четко сформулированы претензии жителей к властям города и строителям (рис. 13).

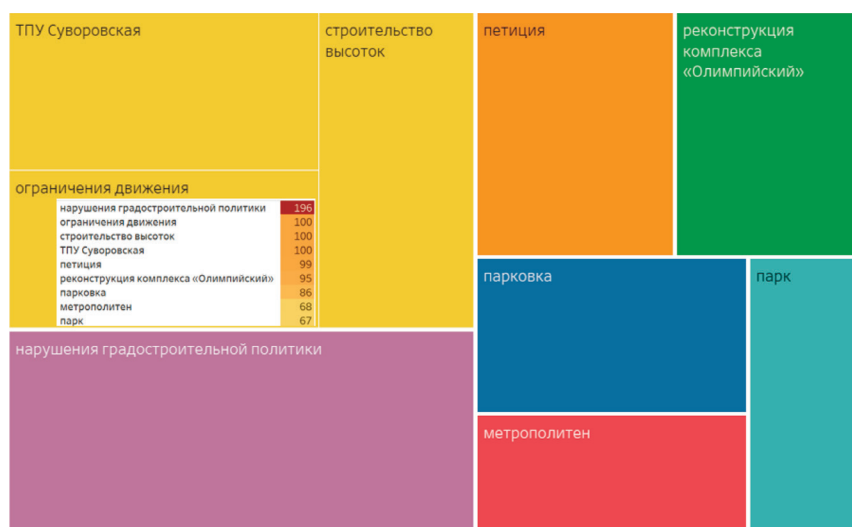


Рис. 13. Ассоциативная сеть стимула ТПУ «Суворовская» имеет (10/551)

Контекст с реакциями.

*Жители Мещанского района против высоток при строительстве ТПУ «Суворовская»*

*Наряду с реконструкцией (сносом) СК Олимпийского (увеличение офисов, пассажирского потока, транспортного коллапса) для жителей подготовили новый КОШМАР в виде застройки площадки ул. Дурова, вл. 3: 1. ЖК высотой 119 м на 50 000 кв.м; 2. Дом жилой по реновации высотой 80 м на 55 000 кв.м; 3. Офисный центр высотой 220 м на 80 000 кв.м. 4. И еще три нежилых дома не менее 200 000 кв.м. Итого около 400 тыс. метров. Второе СИТИ!*

*В результате реализации проекта: полностью будет разрушена биосфера Екатерининского парка и прилегающих традиционных природных мест отдыха жителей; транспортная доступность района для жителей станет критичной; резко вырастет пассажиропоток (в 5–6 раз) на территории района; район перестанет быть местом для полноценного отдыха жителей Мещанского и Тверского районов в выходные дни и нерабочее время.*

Наличие социального стресса в районе строительства ТПУ «Суворовская» осложняется также общим низким уровнем доверия населения властям и средствам массовой информации. Индекс социального благополучия определяется большим количеством ангажированного контента, который выражает официальную точку зрения строителей и городских властей. Между тем у участников конфликта, напротив, вызывает негативную оценку (рис. 14).



Рис. 14. Индексы социального стресса и благополучия

### 3. ЗАКЛЮЧЕНИЕ

Анализ данных показал, что большая часть контента, посвященного ТПУ «Суворовская», носит официальный характер. Посты в социальных медиа являются по большей части репостами и рерайтом официальных сообщений, информации ангажированных ресурсов и не вызывают положительной реакции жителей района.

Максимальное количество сообщений, раскрывающих различные аспекты реализации проекта ТПУ «Суворовская», представлено в микроблогах, некоторое количество — в социальных сетях, незначительное количество — в интернет-СМИ, на тематических порталах, в мессенджерах, блогах, на форумах. Между тем непосредственная реакция пользователей (лайки, комментарии, просмотры) представлена в социальных сетях.

Результат анализа данных показал наличие среднего уровня социального стресса с коэффициентом 16,69 и низкого уровня индекса социального благополучия с коэффициентом 7,65.

Динамика коммуникационных процессов в информационном пространстве вокруг реализации проекта ТПУ «Суворовская» характеризуется эскалацией конфликта между жителями, строителями и властями города.

Анализ данных позволяет говорить о развитии конфликта по мере реализации проекта. На протяжении анализируемого периода можно отметить значительный рост количества сообщений, посвящённых реализации проекта ТПУ «Суворовская». Причем анализ тональности контента демонстрирует рост негативных сообщений и уменьшение позитивных упоминаний. Увеличение нейтральных упоминаний определяется ростом числа официальных материалов, генерируемых ангажированными медийными ресурсами, контент которых вызывает неоднозначное, часто неодобрительное отношение жителей. Подобная ситуация характеризует негативное восприятие пользователями развития событий и эскалацию конфликта.

### **Литература**

1. *Hickok G., Poeppel D.* Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences* 4, 131–138 (2000).
2. *Hickok G.* Functional Anatomy of Speech Perception and Speech Production: Psycholinguistic Implications. *J Psycholinguist Res* 30, 225–235 (2001).
3. *Demuth K.* Perception, production, and individual differences. *Applied Psycholinguistics* 39(4), 735–741 (2018).
4. *Ransom T.G., Dale R., Kreuz R.J., Tollefsen D.* How Do Different Types of Alignment Affect Perceived Entity Status? *J Psycholinguist Res* 48, 961–985 (2019).
5. *Kreyßig N., Krautz A.E.* Lying and perception of lies by bilingual speakers. *Applied Psycholinguistics* 40 (5), 1313–1329 (2019).
6. *VanRullen R.* Perception Science in the Age of Deep Neural Networks. Specialty Grand Challenge Article. *Front. Psychol.*, 02 February 2017.
7. *Mansimov E., Parisotto E., Ba L.J., Salakhutdinov R.* Generating images from captions with attention. *CoRR* 1511.02793 (2015).
8. *Nguyen A., Yosinski J., Bengio Y., Dosovitskiy A., Clune J.* Plug & play generative networks: conditional iterative generation of images in latent space. *arXiv preprint arXiv: 1612.00005* (2016).
9. *Liu H.-M., Tsao F.-M., Li P.* (Eds.) *Speech Perception, Production and Acquisition. Multidisciplinary approaches in Chinese languages.* Springer Singapore, Singapore (2020).
10. *Antons J.-N.* *Neural Correlates of Quality Perception for Complex Speech Signals.* Springer International Publishing, Cham (2015).
11. *Essam B.A., Abdo M.S.* How Do Arab Tweeters Perceive the COVID-19 Pandemic? *J Psycholinguist Res* 1–15 (2020).
12. *Kharlamov A. Pilgun M. (eds.)* *Neuroinformatics and Semantic Representations. Theory and Applications.* Cambridge Scholars Publishing, Newcastle upon Tyne (2020).
13. *White M.D., Marsh E.* Content analysis: a flexible methodology. *Library trends* 1(55), 22–45 (2006).
14. *Krippendorff K.* *Content Analysis. An Introduction to Its Methodology.* 3rd. ed. SAGE Publications, Inc., Los Angeles, United States (2012).

## **REAL-TIME COGNITIVE RESEARCH: A NEURAL NETWORK APPROACH**

***Pilgun M.A., Russian State Social University, Institute of Linguistics,  
RAS, Moscow, Russia, pilgunm@yandex.ru***



The article presents the results of a cognitive study devoted to the analysis of the relationship between residents and city authorities using a neural network approach. The material for the study was data from social networks, microblogging, blogs, instant messengers, forums, reviews, video hosting services, thematic portals, online media, print media and TV related to the construction of the Suvorovsky transport hub in the center of Moscow (RF). To analyze the content of social media, a multimodal approach was used involving neural network technologies, text analysis, sentiment analysis, analysis of word associations and content analysis. The dynamics of communication processes in the information space around the project is characterized by an escalation of the conflict between residents, builders and city authorities. Data analysis suggests further development of the conflict as the project progresses.

• *Cognitive Research* • *Speech Perception* • *Social Media Data* • *Neural Network Technologies*

# Нейродидактика и преподавание иностранных языков и перевода в медиапространстве — на примере немецкого подкаста университета Испании

*Коренева О.Б., PhD, Университет Пабло-де-Олаvide, Факультет Переводчиков, Карретера де Утрера, 1, 41013 Севилья, Испания*

*Лимбах К, PhD, Университет Пабло-де-Олаvide*

*Штендер А., Ph D, Университет Пабло-де-Олаvide*

*Гор М., ДААД, Немецкий академический центр по обмену, Маркграфенштрассе 37, 10117 Берлин, Германия*

Данная работа описывает инновативный проект практического применения нейродидактики в методологии обучения иностранным языкам и переводу в высших учебных заведениях. Нейродидактика также занимается проблемами мотивации обучения [Grein 2021] и опирается на раннее практическое применение приобретенных знаний и навыков — так называемое корпорейзированное обучение — *embodied learning* [Stolz 2015]. Проект состоял в реализации университетского подкаста Radio Wunderbar на немецком языке, где учащиеся факультета переводчиков испанского Университета Пабло-де-Олаvide смогли непосредственно применить свой лингвистический профессиональный опыт в реальных условиях — *being-in-the-real-world*. Участники подкаста сами руководили процессом своего обучения, более эффективно и быстро развивали вербальные и аудитивные навыки на иностранном языке благодаря погружению в реальный контекст и дискурс коммуникации в социальном, цифровом медиапространстве. Данная интерактивная методология продемонстрировала раннее практическое применение усвоенных навыков, что повысило заинтересованность и мотивацию учащихся в обучении иностранным языкам и будущей профессии. Нейродидактический подход к обучению в медиапространстве гарантирует будущим профессионалам готовность к новым вызовам, бросаемым быстро развивающимися мультимодальными речевыми технологиями.

• нейродидактика • интерактивная методология обучения иностранным языкам в медиапространстве • обучение переводчиков • дидактический подкаст





## 1. ВВЕДЕНИЕ

Нейротехнологии и когнитивные науки сегодня все больше набирают силы во всех областях жизни человека. К сожалению, их массовое применение в образовании еще не достигнуто. Последними тенденциями в обучении являются нейропедагогика и теории корпорезированного обучения, или *embodied* или *situated learning* [Stolz 2015], которые очень эффективны в обучении иностранным языкам. Нейродидактика является устойчивой основой обучающего процесса в эмоциональном аспекте, мотивации и функции памяти [Grein 2021].

В области профессионального лингвистического образования стоит понимание устного и письменного дискурса, а также концептуальная компетенция, которые особенно важны для будущих переводчиков. Как известно, контекстуализация и практическое применение дискурса необходимы для его понимания. Традиционным методом для приобретения контекстуализированного словарного запаса с давних пор являлось чтение книг, что все меньше пользуется успехом в цифровую эру. Новые технологии, виртуальная реальность, огромное количество источников, методологий и ресурсов, а также неограниченный поток информации в состоянии коллапсировать восприятие любого и может привести к отчуждению и тем самым к снижению мотивации и интереса к обучению. Кроме того, ритм жизни ускоряется. И если учащиеся высших учебных заведений должны ждать 4–5 лет для применения на практике своих приобретенных знаний, то в случае с будущими переводчиками, например, это негативно отражается на мотивации обучения иностранным языкам и профессиональных материй.

Встает необходимость непосредственного практического применения приобретенных знаний на протяжении обучающего процесса уже на раннем его этапе, привлекая последние достижения в дидактике и самые эффективные технологические ресурсы.

В данной работе описан новый подход к методам обучения иностранным языкам и лингвистическим компетенциям, необходимым для будущих переводчиков. В Университете Пабло-де-Олаvide в Испании была применена нейродидактическая методология, основанная на реальном опыте самих обучающихся немецкому языку. Проект заключался в участии будущих переводчиков в подкасте на немецком языке, где совмещался новый когнитивный, интерактивный подход ситуативного обучения (*embodied/situated learning*) и новых мультимедийных технологий, а конкретно: медиапространство подкаст.

В рамках этого инновационного проекта планировалось развитие гибридной методологии, отличающейся эмоционально заряженным опытом самих учащихся, которые бы лично проводили интервью, создавали диалоги, пели, играли на музыкальных инструментах немецкие мотивы и шутили, создавая реальный медиапродукт в форме программ подкаста. Данное чувство реальности, или *being-in-the-real-world*, должно было помочь им собирать непосредственный практический лингвистический опыт и позволить раннее применение приобретенных знаний

уже на самом базовом уровне. Такой подход должен был в конечном итоге вылиться в повышенную заинтересованность и мотивацию в обучении иностранным языкам и будущей профессии.

Кроме того, методология была направлена на то, чтобы учащиеся с самого начального этапа развили и совершенствовали вербальную и аудитивную способность к иностранному языку. Такие компетенции жизненно необходимы будущим переводчикам и могут и должны практиковаться на самом начальном уровне. Таким образом, основной идеей данного проекта было достижение целостного, холистического процесса обучения переводчиков [Koreneva 2020a]. Конечной целью данного исследования было создание интерактивной методологии эмотивного обучения в реальной ситуации, что превратило бы пассивное обучение (чтение, слушание записей аудио) в активное практическое применение иностранного языка в разговорной речи, сбор практического лингвистического опыта и развитие необходимых навыков, активного дискурса, его запись, оценку, корректировку и воспроизведение. Данный нейродидактический подход с привлечением ресурсов мультимедиа в медиапространстве должен был помочь раннему, практическому и эффективному приобретению языковых знаний и профессиональных навыков учащимися.

## 2. ТЕОРЕТИЧЕСКИЕ ОСНОВЫ

Нейродидактика набирает все больше и больше силы и вместе с другими когнитивными науками вносит практический и полезный вклад в процессы обучения во всех сферах и на всех уровнях. Тенденция процесса обучения отличается все более усиленным применением практического и эмпирического подхода к приобретению знаний. Также современные технологии все больше интегрируются в качестве инструментов обучения, что можно было наблюдать на опыте последних лет в условиях удаленного обучения по вине пандемии ковида-19. В конкретном случае в области профессии устных и письменных переводчиков, чья работа требует особых когнитивных способностей, существует множество исследований, показывающих важность раннего практического применения и усовершенствования профессиональных навыков [Padilla 2002, Limbach 2017; Collados Aís, Krüger und Pradas Macías 2013].

Особенно устным переводчикам важно быстрое и хорошее аудитивное понимание и вербальное воспроизведение дискурса, а также метафорическая компетенция [Danesi 2017j, Stender 2019]. Профессионалам в области устного перевода приходится контролировать *switching costs* рабочих языков, причем под давлением и в короткий отрывок времени. Этими процессами, а также выбором необходимой лингвистической стратегии и удержанием информации заведует рабочая память (*working memory*) и исполнительный контроль (*executive control*). Различные экспериментальные исследования подчеркивают необходимость максимально раннего и интенсивного развития данных механизмов, а также памяти для реализации работы переводчиков [Ibáñez, Macizo & Bajo 2010; Morales et al. 2015 и другие].

Также многочисленные психолингвистические и нейролингвистические исследования указывают на такие основы когнитивной функции человека, как восприятие, действие и симуляция. Речь идет о корпорейзировавшем опыте или обучении (*embodied learning*) [Stolz 2015], где учащийся учит и применяет иностранный



язык естественным образом, погружаясь в симулированную реальную ситуацию и контекст [Danesi 2017]. Данный подход разделяет и нейродидактика, чьи основы опираются на эмоции, мотивацию и раннее практическое применение знаний [Grein 2021].

С недавнего времени в лингвистике существует понятие «вербальность». Хиршфельд и коллеги [2021] говорят о концептуальной и медиальной вербальности. Медиальная вербальность подразумевает звучащий язык, т. е. воспроизведение и звучание языка. Концептуальная вербальность имеет в виду выразительность языка как в письменной, так и в устной форме. Все эти навыки чрезвычайно важны в обучении будущих профессионалов перевода.

Как всеобщее известно, самыми трудными навыками в изучении иностранного языка являются аудитивное понимание и вербальная коммуникация. С растущей глобализацией мира данные навыки даются всё труднее. Как указывает [Kamerer 2019], на сегодняшний день уже невозможно говорить о настоящих «носителях языка», т. к. существует много акцентов и форм применения языка, которые при этом являются эффективными и отлично работают в коммуникации. Также Чучупал [2020] говорит о проблемах распознавания речи механизмами искусственного интеллекта. По этой причине глобализация бросает также вызов переводчикам, которые должны быть готовы столкнуться с бесконечным числом вариантов языка, на котором они работают. Это еще больше усложняет аудитивное восприятие и понимание дискурса и контекста, так необходимых переводчикам. Кроме того, будущим профессионалам придется столкнуться с различными новыми формами коммуникации, в том числе удаленной, и коммуникации в социальном медиaprостранстве.

Другой новой растущей отраслью применения языковых и переводческих знаний и навыков является также сфера медиаторства. Так, все более активные миграционные передвижения людей внутри и вне Европейского сообщества диктуют свои условия на рынке труда и социальных услуг. На 2021 год в Испании проживало 5,38 миллиона официальных мигрантов, что составило 11,34 % всего населения (Statista), а в Германии осело 21,9 миллиона мигрантов (26,7 % населения) (Bundeszentrale für politische Bildung). Данные цифры, а также актуальный кризис беженства гарантируют работу переводчикам и межкультуральным медиаторам на многие годы вперед, но и ставят перед ними новые задачи, где речевые технологии все больше расширяют свое определение.

Социальные службы многих стран Европы все больше нуждаются в экспертах мультилингвальной коммуникации. Как было сказано выше, с недавних пор к лингвистической компетенции переводчиков прибавилась культуральная и медиаторская, которая еще более усложнила процесс перевода и его обучения, а также ставит переводчиков-медиаторов под еще большее психологическое давление со стороны мультикультуральных сообществ [Limbach 2018, Marchioni, 2019, Ramasco Gutiérrez, Giménez Romero & Marchioni, 2020, Koreneva 2020b]. Работая

с немецким и испанским языками, переводчику необходимо знать ситуацию, специфику и предпочтения политики Испании и Германии, а также требования ее лингвистического рынка. Кригель-Шмидт [2012] утверждает, что компетенции медиатора совпадают с интеркультуральными навыками.

Как известно, медиапространство и соцсети являются современным коммуникативным пространством нового поколения [Харламов, Пильгун 2020; Митрофанова 2019], для сосуществования в которых необходимы толерантность, нейтральность и дипломатичность, навыки, так необходимые также и переводчикам [Martín & Abril, 2002; Hale 2010]. Данное исследование следует последним тенденциям на рынке профессиональных услуг перевода и готовит будущих профессионалов к их новой роли социальных, а также и удаленных переводчиков, привлекая последние технологии. Удаленный перевод (*remote interpreting*) подразумевает телефонный перевод, перевод на видеоконференции и другие его формы, когда переводчик не делит одно и то же пространство с клиентом. Новые технологии потворствуют появлению более современных форм работы переводчиков и открывают новые перспективы для его профессионального будущего на рынке труда. Но данное развитие также несет с собой и большую когнитивную нагрузку, дополнительные аудитивные сложности (за неимением визуального контакта или по вине помех), подразумевает умение обращаться с соответствующими техническими средствами [Corpas, Gaber 2020].

Таким образом, встает острая необходимость развития методологии, основанной на гибридном применении новых технологий вкпе с эмпирическим подходом к обучению. Так, серийная аудиопрограмма подкаст — как новая форма коммуникации в сообществе — представляет собой социальное медиапространство, привлекательное и динамичное, идеально подходящее для дидактических и публицистических целей. Подкаст существует уже около десяти лет и отличается всеобщей свободной доступностью с любого устройства (компьютер, смартфон, планшет). К тому же речь идет о разносторонней возможности коммуникации — темы подкаста могут быть разными, молодежными, актуальными. Аудиозаписи программ делаются легко, из любого места, даже в домашних условиях, также с любого устройства, прибегая к бесплатным аудиопрограммам — *Audacity* или *Windows Movie Maker* — и просто вешаются на интернет-страницу в режиме реального времени или для его прослушивания позже. При оптимальных условиях качество записи может приближаться к качеству записи в студии. Для их дидактизации уже существуют разные формы. Так, например, известны подкасты для прослушивания медленно объявляемых новостей на немецком языке в *Deutsche Welle* (*langsam gesprochene Nachrichten*: <https://www.dw.com/de/nachrichten/s-8030>).

Многие авторы посвятили время изучению преимуществ и недостатков обучения с помощью подкастов [см. например: Orlando 2011, Stork 2012]. Положительными пунктами подготовки материала для подкаста могут выступить самостоятельная тренировка и запись в собственном ритме в любом месте, с последующим прослушиванием, контролем, корректировкой и совершенствованием произношения. По выбору учащегося и преподавателя материал может готовиться индивидуально или в группе. По сравнению с работой в классе, где необходимо подстраиваться под общий ритм и уровень, подготовка аудиоматериала может проходить в качестве домашнего задания на свободную или заданную тему, развивая креативность и внося творческий аспект в обучающий процесс. Речь идет о *mobil-learning* (*m-learning*) или мобильном обучении при помощи электронных устройств, независимо



от места нахождения. При этом Шмидт [2010] предупреждает, что подобная дидактика требует ее разработки, новых качественных форматов тренировок и материалов, которые бы мотивировали учащихся. Одна только техника не заменит преподавателя и не имеет дидактический потенциал [Stork 2012].

Таким образом, становится очевидно, что вербальная, аудитивная и социальные компетенции должны идти рука об руку в процессе обучения такой сопровождаемой сложными когнитивными и лингвистическими процессами профессии, как профессия переводчика, представляя собой целостный холистический подход [Koreneva 2020a].

### 3. МЕТОДОЛОГИЯ ПОДГОТОВКИ ПОДКАСТА *RADIO WUNDERBAR* НА НЕМЕЦКОМ ЯЗЫКЕ

Надо сказать, что уже повсеместно существуют курсы подготовки подкастов для разных целей. Однако, насколько нам известно, существует мало опыта подготовки подкаста для дидактических целей иностранных языков и профессиональных навыков. В данной работе описывается подход и процесс подготовки подкаста *Radio Wunderbar* для обучения немецкому языку и устному переводу.

**Инструмент:** в настоящее время можно рассчитывать и бесплатно пользоваться программой *Audacity 3.1.1*, которая находится в свободном распоряжении пользователей Интернета. Этот простой в своем использовании редактор звуков напоминает программу *Microsoft Word*, где вместо текстов записываются и обрабатываются цифровые аудиофайлы по тому же принципу. Самостоятельно записанные студентами (индивидуально или в группе) при помощи смартфона, планшет или компьютера лично или удаленно (Zoom, Google Meet, Hangout, Skype, blackboard, etc.) аудиофайлы могут сохраняться и обрабатываться учащимися и координаторами программы в формате .m3. Файлы были посланы по электронной почте, WhatsApp или вывешены учащимися на обучающей университетской платформе Aula Virtual, blackboard collaborate соответствующего предмета.

**Длительность:** данное исследование базируется на опыте подготовки программ для подкаста с октября по декабрь и февраля по март 2020–21 учебного года — всего восемь программ по 59 минут каждая. Программа выпускалась в последний четверг каждого месяца в прямом эфире с 13–14 часов, а затем ставилась в фонотеку, для последующего прослушивания.

**Участники:** в ней приняли участие студенты факультета переводчиков и международных отношений со специализацией в переводе и немецком языке (основной рабочий язык). В частности, речь идет о студентах с первого по четвертый курс и предметах: немецкий язык 1 и 2, социолингвистика/культура и общество, введение в практику перевода, двусторонний устный перевод, всего 125 участников.

**Структура:** по причине различных уровней владения немецким языком учащихся 1–4-го курсов и различными предметами, которые составляли тематику подкаста, программа повышала уровень сложности с ее начала и до конца, где в начале выступали первокурсники с легкими диалогами, а заканчивалась программа докладами и презентациями, а также интервью с симультанным устным переводом старшекурсников.

**Материал:** материалом для подкаста *Radio Wunderbar* служили темы лекций и различные отдельные задания, интервью, служащие для развития различных компетенций: дискурсивной, лексической, тематической, культуральной и т. д. Кроме того, координаторы стремились включить темы актуальности: национальные праздники (Рождество, Октоберфест в Мюнхене, карнавал в Кельне, пасхальные ходы в Севилье) и осветить последние и знаменательные события, имеющие место в Испании и Германии. Также приветствовалась свободная инициатива и применение к немецкому творческих способностей студентов (пение, игра на музыкальных инструментах), а также повествование о личном опыте с немецким языком (личный подход к обучению, советы по усвоению словарного запаса, подводные камни, а также поездки по студенческому обмену в немецкоговорящие страны, работа за границей на летних каникулах и другое). Предпочтение отдавалось темам, волнующим молодежь, и интересным слушателям, и обучающимся немецкому языку.

**Форма работы:** приветствовалась командная работа и работа в группе, что служит развитию социальной компетенции и способности работать в команде.

**Формат:** программа подкаста *Radio Wunderbar* записывалась в формате .mp3 stereo и размещалась на сайте университетского радио *RadiOlavide* <https://radiolavide.org/categoria-audios/radio-wunderbar>. Язык программы был в основном немецкий, хотя в некоторых ее частях прибегался к билингвальности (повторение текста на испанском языке) для упрощения понимания тем начинающими изучать немецкий язык. Отдельные записанные студентами аудиофайлы представляли собой короткие диалоги (1 мин), монологи (2–3 мин), презентации (5 мин), музыкальную аранжировку, даже пение немецких песен, а также интервью и их симультанный перевод (8–10 мин). По инструкции презентация должна быть интерактивной и подразумевать прямое обращение к слушателю, что помогало усилить эффект реальности ситуации. Темы были материалами уроков по предметам на немецком языке, чьи преподаватели участвовали в проекте, или свободно выбранные самими учащимися. В основном речь шла о разговорной речи, практической культурной и языковой информации.

Координатор программы (первый автор) собирала аудиофайлы, полученные доцентами-участниками проекта (автор 2–4) от их учеников, и совмещала их в одну запись, проводя соответствующую звуковую редакцию, уравнивание и чистку звукового фона, вставку специальных эффектов и музыкальных заставок, укорачивая ненужные паузы и вырезая лишние элементы. В случае специальных эффектов и музыки необходимо указать на обязательное соблюдение прав автора и рекомендуется использовать ресурсы свободных от прав автора или материалы по предварительной договоренности с автором или с разрешения родителей (как в случае с материалами, предоставленными несовершеннолетними). Учащимся были даны соответствующие инструкции по соблюдению всех правил.



#### 4. РЕЗУЛЬТАТЫ ПРАКТИЧЕСКОГО ПРИМЕНЕНИЯ РЕЧЕВЫХ ТЕХНОЛОГИЙ В ПОДКАСТЕ *RADIO WUNDERBAR*

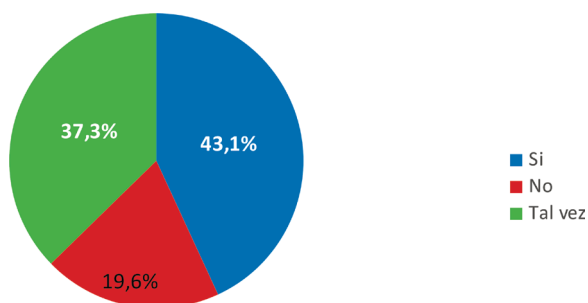
Результатом проекта раннего практического применения языковых и профессиональных навыков будущих переводчиков в медиапространстве стал развлекательный обучающий медиапродукт: подкаст *Radio Wunderbar* с восемью программами на немецком языке. В аудиозаписи продолжительностью в 59 мин уровень содержания и трудности шел по восходящей к концу программного времени, начиная с простых диалогов первокурсников, заканчивая симультанным устным переводом интервью учащихся 4-го курса. Программы распределились с октября по июнь, тогда как январь выпадал на сессию, и в это время подкаст не записывался. В результате размещения подкаста на сайте университетского радио *RadiOlavide* было зарегистрировано среднее количество прослушиваний программ — 188 слушателей на программу, что составило очень высокий рейтинг университетского радио, который обычно достигал максимально 40–50 прослушиваний на программу. Первая же вышедшая в эфир октябрьская программа подкаста собрала 509 слушателей, что побило все рейтинги прослушиваний *RadiOlavide* с начала его существования (2016). Всего в программе приняли участие 125 учащихся. Программа состояла из четырёх частей, соответствующих четырём предметам, чьи преподаватели координировали проект. Первая и вторая части программы содержали также билингвальные дискурсы для облегчения понимания на этих базовых уровнях. Надо подчеркнуть творческий подход и юмористические нотки, которые благодаря участникам подкаста просматривались в каждой его программе, тогда как заставки современной молодежной немецкой музыки и специальные эффекты придали ценности развлекательной обучающей программе.

Участие в интерактивном проекте обучения в форме подкаста сопровождалось повышением интереса учащихся и их уверенности в себе на уроках и при выполнении заданий на немецком языке. Ранняя осознанность необходимости обучения определенным навыкам и компетенциям, а также понимание целесообразности поставленных перед будущими переводчиками задач положительно отразились на мотивации в процессе обучения (и преподавания) и вылились в более высокие оценки по соответствующим предметам. Данный подход оказался эффективным для подготовки студентов начиная с первого курса для успешного прохождения практики устного перевода в более старших курсах и усовершенствования общего владения дискурсом в любой области.

В конце учебного года был также проведен дополнительный опрос участников подкаста *Radio Wunderbar*, результаты которого можно увидеть ниже.

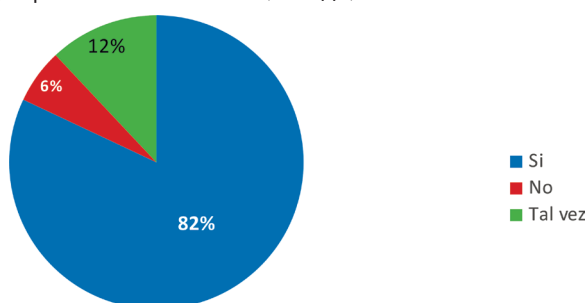
Так, на вопрос: «Стал ли меньше страх перед публичным выступлением на немецком языке у его участников?» 44 % ответили — да, 38 % ответили, что возможно (фигура 1).

Также в конце опроса была предложена возможность поделиться своим мнением о подкасте, в ответ на что были получены самые положительные отзывы, в числе которых говорилось о новизне и разнообразии



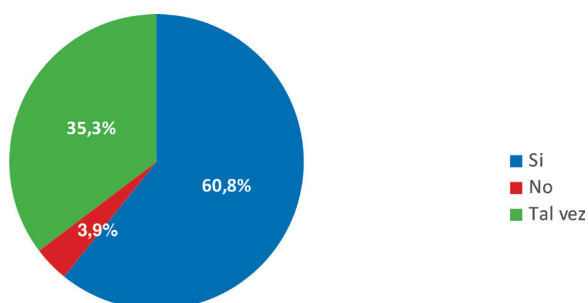
Фигура 1

Фигура 2 показывает ответы на вопрос: «Думают ли они, что участие в подкасте способствует работе в команде?» 82 % с уверенностью ответили, что да, и только 12 % — может быть.



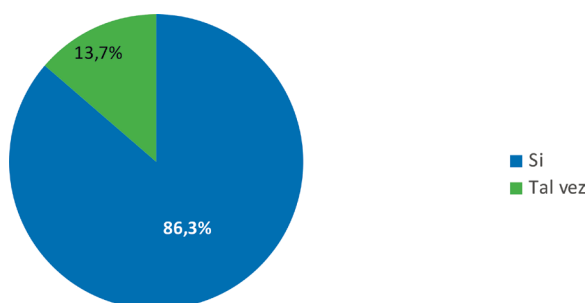
Фигура 2

На вопрос: «Считаешь ли ты, что после участия в подкасте твои дискурсивные способности и устное выражение на немецком улучшилось?» 61 % ответили, что да, 36 % — возможно (фигура 3).



Фигура 3

На вопрос: «Понравился ли вам опыт и рекомендовали бы вы его сокурсникам?» 87 % ответили, что да, только 14 % — может быть, и никто не сказал, что нет (фигура 4).



Фигура 4





в обучении благодаря участию в подкасте, а также выразалось желание углубить навыки пользования ресурсами мультимедиа.

## 5. ДИСКУССИЯ И ВЫВОДЫ

Учитывая результаты проекта и опроса участников, а также рейтинг университетского немецкого подкаста *Radio Wunderbar* для раннего обучения и практики немецкого языка и перевода, можно утверждать, что интерактивный мультимедийный подход к обучению иностранным языкам и профессиональным навыкам является перспективным и эффективным направлением в дидактике.

Подготавливая свои короткие доклады на немецком языке для подкаста, учащиеся работали в основном в команде, выступали в активной роли журналистов, берущих интервью, преподавателей немецкого языка, давая советы и делаясь собственным опытом обучения, вели диалоги, модерировали в реальном медиапространстве. Речь идет о самостоятельной интерактивной деятельности, несущей в себе эмоциональную окраску и предполагающую корпорезированный, когнитивный опыт в реальной ситуации.

Кроме того, данный мультимодальный подход к обучению позволяет учащимся уже на раннем периоде примерить на себя роль переводчика, проверить свою социальную и коммуникативную компетентность, что помогает достигнуть прогресса и повышает интерес к учебе и мотивацию.

Погружение в реальную ситуацию в медиапространстве немецкого подкаста гарантирует ситуативное обучение и представляет собой методологию ранней практической профессиональной подготовки будущих переводчиков. Она позволяет учащимся получить ранний лингвистический опыт в реальных условиях — *being-in-the-real-world* [Stolz 2015] — на самом начальном уровне. Такой подход к обучению повышает сознание практической необходимости усвоенных навыков и тем самым улучшает заинтересованность и мотивацию учащихся в обучении иностранным языкам и их будущей профессии [Grein 2021], а также их уверенность в себе при реализации публичного дискурса. Кроме того, участвуя в подкасте на немецком языке, учащиеся с самого начального этапа играют ведущую роль в процессе своего обучения, более эффективно и быстро развивают вербальные и аудитивные навыки на иностранном языке благодаря погружению в реальный контекст и дискурс коммуникации в социальном, цифровом медиапространстве.

Таким образом, данный инновационный проект разработки методологии преподавания немецкого языка для будущих переводчиков совместил новый когнитивный, интерактивный подход ситуативного обучения (*embodied/situated learning*) и новые технологии.

Наряду с развитием вербальной и аудитивной способностей на иностранном языке также преследовалось приобретение повышенной социальной компетентности, необходимой будущему переводчику при совмещении

его функций с медиаторством в мультикультуральном сообществе. Знакомство с новыми технологиями и использование программ мультимедиа позволило учащимся открыть для себя новые перспективы будущего и подготовиться к вызовам, которые бросают им новые формы работы, такие как удаленный перевод и другие. Все эти компетенции являются жизненно необходимыми для будущих переводчиков и могут и должны развиваться уже на самом начальном уровне и совершенствоваться в дальнейшем, выливаясь в холистический процесс обучения профессионалов высокого уровня [Koreneva 2020a].

Данное исследование доказывает эффективность, целесообразность и необходимость имперических мультимодальных методов раннего практического обучения речевым технологиям, основанным на нейродидактике и имеющим место в медиапространстве.

Кроме того, можно смело утверждать, что данная предложенная методология может быть внедрена также и в другие области обучения, в том числе она также приемлема в качестве дистанционного обучения. Развитие социальных компетенций, раннее практическое применение профессиональных навыков, способность работать в команде, создавая общий продукт, и владение дискурсом важны в любой сфере деятельности человека. Быстрое развитие новых речевых и мультимедийных технологий и социальные изменения в обществе заставляют дидактику идти с ними в ногу и предоставляют новое поле деятельности на благо образования и мирного сосуществования мультикультуральных сообществ.

## Литература

1. Митрофанова, И.И. (2019). Клиповое мышление: реальность и перспективы. Речевые технологии 2019.
2. Чучупал, В.Я. (2020). Нейросетевые модели языка для систем распознавания речи. Речевые технологии, 1–2–2020.
3. Харламов, А.А. (2019). Семантика текста как модель ситуации. Речевые технологии 2019-1, стр. 14.
4. Харламов А.А. (2019). Нейросетевой подход к распознаванию ситуации по тексту. Речевые технологии 2019-1.
5. Kharlamov, A.A., Pilgun, M.A. (2020). Perception and Effectiveness of Content in Social Networks: Russian Case, Speech Technology 2–2020.
6. Cramerer, R. (2019). Aufzeichnung zum Thema "Der neue Begleitband zum GER des Europarats: Auswirkungen auf den Fremdsprachenunterricht?" vom 19.11.2019, Available: <https://www.youtube.com/watch?v=b2vhpMUn2ns>
7. Collados Aís, Á., Krüger, E., Pradas Macías, M. (2013). Selbstlernkurs Gesprächsdolmetschen (Handbuch und interaktive DVD), Granada: Comares.
8. Corpas Pastor, G., Gáber, M. (2020). "Remote Interpreting in Public Service Settings: Technology, Perceptions and Practice". SKASE Journal for Translation and Interpretation, 13 (2), 58–68. ISSN: 1336–7811.
9. Danesi, M. (2017): Conceptual Fluency Theory and the Teaching of Foreign Languages (Education in a Competitive and Globalizing World: Languages and Linguistics). Ph.D. Nova Science Publishers, Inc (23. Juni 2017).
10. Grein, Marion (2021). Flexibles Lernen und Interaktion aus neurobiologischer Perspektive, Seminar für Germanistenverband Andalusiens, Verlag Hueber/Fr. Dr. Marion Grein, Universität Mainz. Neurodidaktik. <https://www.germanistas.com/cursos/>



11. *Hirschfeld, U., Rösler, D., Schramm, K.* (2021). Facetten der Mündlichkeit im DaF-Unterricht. Zur Einführung in den Themenschwerpunkt, ed. Erich Schmidt Verlag GmbH & Co. KG, Berlin 2021 — (esv-campus.de) — 27.04.2021, Deutsch als Fremdsprache 3/2016.
12. *Ibáñez, A.J., Macizo, P., M.T. Bajo* (2010): "Language Access and language selection in professional translators". *Acta Psychologica*, volume 135, issue 2, Elsevier, 257–266. <https://doi.org/10.1016/j.actpsy.2010.07.009>
13. *Koreneva, Olga* (2020a). Embodied learning of the German language: teaching methodology for translators and interpreters. *mAGAZin*, ISSN-e 1136-677X, N<sup>o</sup>. 28, 2020, págs. 59–67.
14. *Koreneva, Olga* (2020b). Interpretación como mediación lingüístico-cultural para proteger los derechos de la mujer migrante. *Derechos Humanos y globalización*. Dykinson, available: <https://www.torrossa.com/en/resources/an/4997922>, P. 701–717
15. *Kriegel-Schmidt, K.* (2012). Interkulturelle Mediation: Plädoyer für ein Perspektiven-reflexives Modell, LIT Verlag Münster, p. 70–72.
16. *Hale, S. B.* (2010). La interpretación comunitaria, la interpretación en los sectores jurídico, sanitario y social. Granada: Comares.
17. *Limbach, C.* (2018). El reto intercultural de las actividades tándem en la combinación lingüística alemán — español. In J.M.Tejedor-Cabrera, J.J. Martos Ramos & L. Trapassi (Eds.), *Aplicaciones de la metodología Tándem en la formación universitaria* (pp. 211–222). Peter Lang. <https://doi.org/10.3726/b14735>.
18. *Limbach, C.* (2017). Ein didaktischer Vorschlag für das Unterrichten von bilateralem Dolmetschen in der Sprachkombination Spanisch-Deutsch. *Glottodidactica*. An International Journal of Applied Linguistics, XLIV (2), 97–113. Available: <https://doi.org/10.14746/gl.2017.44.2.06>
19. *Marchioni, M.* (2019). La intervención comunitaria como instrumento de cambio. En: Zarco J, Ramasco M., Pedráz A., Palmar AM. *Investigación cualitativa en Salud*. 1<sup>a</sup> ed. Madrid: Centro de investigaciones sociológicas, p. 345.
20. *Martín, A. & Abril, I.* (2002). Los límites difusos del papel del intérprete social en España. En Valero Garcés, C. y G. Mancho Barés (Eds.). *Traducción e interpretación en los servicios públicos: nuevas necesidades para nuevas realidades*. Community interpreting and translating: new needs for new realities. Alcalá de Henares: Universidad de Alcalá, pp. 57–63.
21. *Morales, J., Padilla, F., Gómez-Ariza, C.J., Bajo, M.T.* (2015): "Simultaneous interpretation selectively influences working memory and attentional networks". *Acta Psychol (Amst.)* 2015, Feb.155, 82–91. doi: 10.1016/j.actpsy.2014.12.004
22. *Orlando, D.* (2011). Didaktische Vorteile von Podcastproduktionen. *Core*. RJIS, Vol. 19 (2). Core.ac.uk.
23. *Padilla Adamúz, F.M.* (2002): Memoria de trabajo y funciones de control en la interpretación de lenguas. Tesis doctoral. Universidad de Granada [2002].
24. *Ramasco Gutiérrez, M., Giménez Romero, C. & Marchioni, M.* 2020. Una década trabajando por la mejora de la convivencia: El Proyecto de Intervención Comunitaria Intercultural. *Revista Madrileña de Salud Pública*. 3, 9 (mar. 2020), 1–7. DOI:<https://doi.org/10.36300/remasp.2020.052>.
25. *Schmidt, Torben.* Multimediale Lernumgebungen für das Fremdsprachenlernen. In: Hallet, Wolfgang; Königs, Frank G. (Hrsg.): *Handbuch Fremdsprachendidaktik*. Seelze-Velber: Kallmeyer, 2010, 280–284.
26. *Stender, A.* (2020). "La metáfora de la crisis financiera como recurso para la enseñanza del lenguaje especializado económico-financiero alemán y español". *Entreculturas* Número 10. *Revista de Traducción y Comunicación Intercultural*, 283–294. Available: <https://revistas.uma.es/index.php/revtracom/article/view/12565>

27. Stolz, Steven A. (2015): "Embodied Learning, Educational Philosophy and Theory: Incorporating ACCESS", *Educational Philosophy and Theory* 47(5), 474–487, DOI: 10.1080/00131857.2013.879694
28. Stork, Anje (2012). Podcasts im Fremdsprachenunterricht — ein Überblick. *Allgemeine Beiträge*. 3–16, Info DaF 1.
29. Bundeszentrale für politische Bildung: <https://www.bpb.de/nachschlagen/zahlen-und-fakten/soziale-situation-in-deutschland/61646/migrationshintergrund>
30. Deutsche Welle (langsam gesprochenе Nachrichten: <https://www.dw.com/de/nachrichten/s-8030>).
31. Radio Wunderbar. RadiOlavide, Universidad Pablo de Olavide. <https://radiolavide.org/programa/radio-wunderbar>
32. Statista: <https://es.statista.com/estadisticas/472404/poblacion-de-espana-por-nacionalidad/>

## **NEURODIDACTICS AND TEACHING OF FOREIGN LANGUAGES AND INTERPRETING IN MEDIA SPACE — ON THE EXAMPLE OF A UNIVERSITY GERMAN PODCAST IN SPAIN**

**Koreneva Olga B.**, PhD, *University Pablo de Olavide, Department  
Translation & Interpreting, Ctra. de Utrera, 1, 41013 Sevilla — Spain*  
*quovadis36@gmail.com*

**Stender Alice**, PhD, *University Pablo de Olavide, idem*  
*aste@upo.es*

**Gohr Marike.**, *Deutscher Akademischer Austauschdienst- DAAD Im  
Wissenschafts Forum am Gendarmenmarkt  
Markgrafenstraße 37, 10117 Berlin*  
*marikegohr@zedat.fu-berlin.de*

This paper describes an innovative project for the practical application of neurodidactics in the methodology of training in foreign languages and translation in higher educational institutions. Neurodidactics is also engaged in the problems of learning motivation [Grein 2021] and relies on the early practical application of acquired knowledge and skills — embodied learning [Stolz 2015]. The project consisted in the elaboration of university podcast in German Radio Wunderbar, where students of the Faculty of Translators of the Spanish University Pablo de Olavide were able to directly apply their linguistic professional skills and to get the being-in-the-real-world-experience. Podcast participants themselves led their learning process, more efficiently and quickly developed verbal and audible skills in a foreign language due to the immersion in the real context and discourse of communication in social, digital media space. This interactive methodology has demonstrated, that the early practical application of assimilated skills has increased the interest and motivation of students in learning foreign languages and the future profession. Neurodidactic approach to teaching in media space guarantees future professional readiness for new challenges thrown by rapidly developing multimodal speech technologies.

• *neurodidactics* • *interactive methodology of training in foreign languages in media space* • *training of translators* • *didactic podcast*



# Речевые технологии развития общения у детей с отсутствием речи

*Ана Медина Регера, PhD, Университет Пабло де Олавиде, кафедра письменного и устного перевода, Ctra. de Utrera, 1, 41013 Севилья - Испания  
anamedina@upo.es*

В данной работе представлены некоторые речевые и коммуникативные технологии, которые меняют качество жизни детей с отсутствием речи, таких, как дети с церебральным параличом и сопутствующими заболеваниями. В частности, мы обращаемся к технологическим достижениям, связанным с программным обеспечением аугментативных и альтернативных систем связи (SAAC), которые позволяют улучшить лингвистическую конфигурацию пиктографических систем. Кроме того, были усовершенствованы способы доступа к этим системам, такие как считыватели глаз и интерфейсы «мозг-компьютер». Из обеих групп мы укажем аспекты использования технологий искусственного интеллекта, ответственных за непосредственное улучшение выразительных возможностей пользователей SAAC: с одной стороны, сочетание пиктограмм и лингвистического предсказания на виртуальных клавиатурах; с другой — технологии искусственного интеллекта, которые позволяют улучшать математические модели доступа, с использованием компьютера. Дополнение и развитие темы будет продолжено в исследовании по проекту «Carases de Comunicar», финансируемом Испанским фондом науки и технологий Министерства науки и инноваций Испании.

• аугментативные и альтернативные системы • дети с отсутствием речи • технологии искусственного интеллекта

## ВВЕДЕНИЕ

Речевые технологии охватывают самые разные области работы и применения. Можно сказать, что это набор методов, которые позволяют взаимодействовать людям и машинам с участием голоса и применением языковых технологий, которые также включают автоматическую обработку письменного языка (саммаризацию, генерирование), например автоматическую речь, распознавание, преобразование текста в речь (синтез, см. Taylor 2009) и диалоговые системы, способные общаться с пользователями-людьми. В области здравоохранения, в клиническом применении, выделяется голосовая биометрия, которая анализирует голоса

людей, выявляя характеристики своеобразных голосов или определённых существующих или потенциальных речевых и языковых расстройств (Fernández García, 2021). Эти технологии сочетают в себе методы обработки сигналов, искусственный интеллект и компьютерную лингвистику. Искусственные нейронные сети и глубокое обучение (Graves, Mohamed, and Hinton, 2013) быстро улучшают вышеперечисленные подходы. Так обстоит дело с программным обеспечением для дополнительной и альтернативной связи (AAC), которое имеет всё более совершенные модули синтеза речи. Средства голосового вывода (VOCA, также SGD: устройства, генерирующие речь) представляют собой высокотехнологичные устройства AAC, которые воспроизводят речь для человека, который не может общаться устно из-за болезни или инвалидности. Устройства воспроизводят оцифрованный или синтетический голос. В случае с оцифрованным голосом запись предварительно записана и воспроизводится, и используется в ряде случаев в низкотехнологичных устройствах (простых коммуникаторах типа Smooth Talker, BigMack или Step by Step). Синтезированные голоса, с другой стороны, основаны на технологиях преобразования текста в речь и заменяют естественный голос человека наиболее естественным образом.

Для испанского выделяются голосовые технологии Acapela, Loquendo или Microsoft. AAC — это одновременно область исследований и набор клинических и образовательных стратегий и практик (Miranda, 2013), которые компенсируют отсутствие естественной речи другими механизмами, такими как жесты, символы и технологические устройства. AAC используется, когда речи или языка недостаточно для удовлетворения потребностей людей в устном общении (Beukelmann & Light, 2020). Пиктографические и письменные системы заменяют естественную речь, а также поддерживают развитие речи у детей с трудностями в понимании и эмоциональной регуляции. Хотя AAC традиционно использовался у людей с физическими ограничениями без связанных с ними умственных нарушений, в настоящее время известно, что он полезен для всех с серьёзными коммуникативными трудностями, включая людей с умственной отсталостью любого диагноза (Schlosser, 2003). Эффективность была показана у людей с синдромом Ангельмана., TEA<sup>1</sup> (Schlosser & Blischak, 2001) синдром Дауна и другие.

Речевые технологии для людей с отсутствием речи в Испании представлены очень мало, и вряд ли найдутся терапевты, неврологи, логопеды или учителя, которые знают, как их внедрить. Усилия Барселонского университета выделяются тем, что соответствующий центр взаимодействует с факультетом психологии, на котором есть подразделение под названием UTAC<sup>2</sup> для Каталонии, а также благодаря усилиям Министерства социальных дел через CEAPAT-Imserso, у которого есть база данных с информацией о САА и вспомогательных продуктах<sup>3</sup>. Из недавних разработок выделяется ассоциация AlfaSAAC, которая создала информационный веб-портал с возможностями обучения и консультаций, использования материалов, библиографии и т. д.

Университет Пабло де Олавиде разработал проект «Carases de Comunicac», финансируемый Испанским фондом науки и технологий, в рамках проекта по распространению и передаче знаний для повышения образовательного уровня населения в области науки и техники. Проект состоит из двенадцати видеинтервью с пользователями

<sup>1</sup> Véase en ruso <https://www.rus-aac.ru/post/альтернативная-и-дополнительная-коммуникация-в-работе-с-детьми-с-рас>

<sup>2</sup> [www.utac.cat](http://www.utac.cat)

<sup>3</sup> Véase [https://ceapat.imserso.es/ceapat\\_01/redes/com\\_aa/index.htm](https://ceapat.imserso.es/ceapat_01/redes/com_aa/index.htm)

и экспертами как из Испании, так и из-за рубежа, а также заключительного круглого стола. С помощью вопросов и ответов интервьюируемые рассказывают о своем опыте из различных областей: собственный опыт (Брэдли Хевен и Кэтрин Лемлер), методология (Майкл Кларк и Йенс Бениш), восприятие читателей (Хуан Хосе Берналь и Ройден Джеймс), особенности мозга и компьютерный интерфейс (Хосе Андрес Гонсалес и Виктор Мартинес), ассоциации и фонды (Морис Гринберг), синтез голоса (Инмакулада Эрнаес), дизайн пиктограмм (Серхио Палао и Аннет Китцингер), предпринимательство в области вспомогательных технологий (Хавьер Монтанер). Все материалы находятся в открытом доступе на сайте проекта<sup>4</sup>.

В данной работе мы освещаем новейшие технологии, которые оказывают реальное влияние на аугментативное общение в детстве. Во-первых, мы остановимся на компьютерных приложениях и технических критериях, позволяющих оценить одни варианты по сравнению с другими; во-вторых, ридеры как средство общения и доступа к компьютеру; в-третьих, некоторые будущие приложения взаимодействия человека и машины и, наконец, последние методологические достижения, направленные на развитие детской речи.

## 1. ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ И ПРИЛОЖЕНИЯ

Бумажные коммуникационные доски (см. рисунок 1) постепенно заменяются современным коммуникационным программным обеспечением, содержащим полные пиктографические словари, доступ к клавиатурам с лингвистическим предсказанием и многочисленными вариантами доступа к ним.



Рис. 1. SAAC, подготовленный непосредственно на бумаге без технической поддержки<sup>5</sup>

<sup>4</sup> La web del proyecto es <https://www.upo.es/investiga/capacesdecomunicar/>

<sup>5</sup> Fuente de la imagen: Creena. Centro de Recursos de Educación Especial de Navarra. <https://creena.educacion.navarra.es/web/necesidades-educativas-especiales/equipo-motoricos/respuesta-educativa-motoricos/saac/8853-2/>



Рис. 2. Коммуникатор, распечатанный с помощью программного обеспечения CAA Grid3

Печатная версия по-прежнему необходима для повседневных ситуаций, в которых нет возможности использовать технику (в ванной, бассейне, во время сеанса физиотерапии или в постели, а также в случае поломки), но теперь мы рассматриваем печатную версию как вторичную версию технологической, то есть содержимое электронного коммуникатора печатается (обе версии одинаковы в случае пиктографических систем). На рис. 2 показана печатная версия программного обеспечения для коммуникации. В то время как для данного ПО традиционно требовалась операционная система Windows, теперь существует мобильный формат в так называемом конвертируемом виде или «два в одном» (то есть ноутбуки в формате планшета, такие как Microsoft Surface), в последние годы его также можно использовать в операционной системе IOS, особенно в iPad, а также в приложениях Android. В Испании для людей, не обладающих навыками компьютерной грамотности и нуждающихся в изучении символической коммуникации, комбинируется коммерческое ПО, особенно Grid3, Proloquo2Go, TDSnap и Verbo, с другим бесплатным ПО, таким как Plaphoons, Jocomunico и LetmeTalk, Symbolotalk. Между ними существуют заметные различия, поэтому некоторые авторы говорят



о таких критериях, как надёжность системы связи. Существует консенсус в отношении того, что система является надёжной, если она позволяет генерировать все коммуникативные функции, включает все классы грамматических и лексических слов, когда она имеет доступ к виртуальной клавиатуре, позволяет вводить формулы вежливости и ритуальные формулы и, самое главное, когда позволяет моторная автоматизация (motor automation), то есть размещение пиктограмм в ячейках, которые всегда находятся в одном и том же месте, и навигация между экранами стабильная, то есть с маршрутами, которые пользователь может научиться объединять, чтобы составлять предложения с суммой нескольких ячеек (Beukelman and Light, 2020). Это большое отличие от системы PECS, основанной на обмене пиктограммами без фиксированного положения и без технической поддержки. PECS была разработана в США в 1985 году Бонди и Фростом для учащихся с РАС, и сегодня предполагается, что многие учащиеся могут перейти с этой системы на устройство, генерирующее голос, о чём можно прочитать на сайте испанской компании, которая продаёт данную систему.<sup>6</sup>

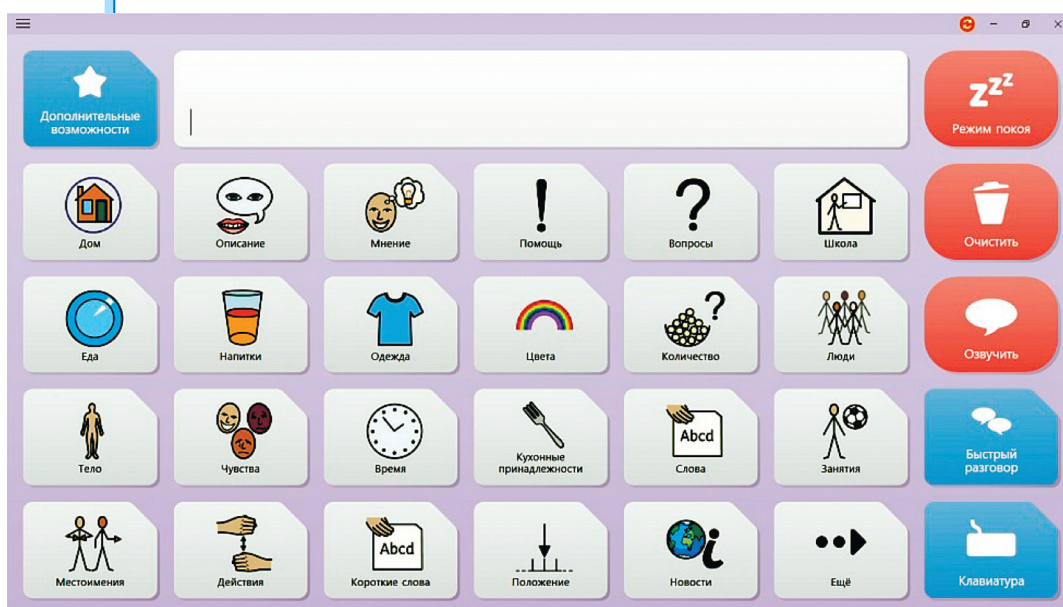


Рис. 4. Один из вариантов главного экрана Grid3, основанный на символах виджета

Настройки для разных приложений также различаются в зависимости от того, как люди получают доступ к устройству. Одной из новейших характеристик компьютера является использование грамматического модуля: пользователь после нажатия на личное местоимение автоматически получает возможность спрягать пиктограмму глагола. Спряжение и времена глаголов, множественное или единственное число существительных и род прилагательных являются наиболее полезными морфосинтаксическими элементами для испанского языка.

<sup>6</sup> <https://pecs-spain.com/el-sistema-de-comunicacion-por-el-intercambio-de-imagenes-pecs/>

Синтаксические правила обеспечивают более естественный расширенный ввод в средствах выражения пользователей и их собеседников.

Наконец, одним из критериев надёжного коммуникационного программного обеспечения является то, что оно может со временем увеличивать словарный запас. Для учащихся с двигательными нарушениями важен размер ячеек (первый ряд изображения 5), помимо моторной автоматизации (чтобы пиктограммы оставались в одном и том же положении). Учащиеся с когнитивными нарушениями получают доступ на ощупь (пальцем), наилучшей формой роста является та, которая позволяет открывать новые клетки по мере изучения словарного запаса (вторая строка изображения 5).



Рис. 5. Две разные организации расширения словарного запаса в программном обеспечении TDSnap

Результаты некоторых приложений представлены в следующей таблице.

Таблица 1

### Сравнение некоторых программ САА

	Grid3	Proloquo2Go	Plaphoons	LetMe Talk	TDSnap
Включает словарный запас, организованный в семантические категории	Да	Да	Нет	Нет	Да
Возможность скрыть ячейки	Да	Да	Нет	Нет	Да
Модуль спряжения и грамматики	Да	Да	Нет	Нет	Да
Клавиатура	Да	Да	Да	Нет	Да
Сенсорный доступ	Да	Да	Да	Да	Да
Переключить доступ	Да	Да	Да	Да	Да
Взгляд - доступ	Да	Нет	Да	Нет	Да
Операционная система Windows	Да	Нет	Да	Нет	Да
Операционная система iOS/Mac	Нет	Да	Да	Да	Нет
Система Android	Нет	Нет	Да	Да	Нет
Бесплатное ПО	Нет	Нет	Да	Да	Нет

## 2. ОБЩЕНИЕ С ЧИТАТЕЛЯМИ ГЛАЗАМИ

Среди вариантов доступа к высокотехнологичным устройствам выделяется разработка айтрекеров (также называемых ай-ридерами или глазами-мышками). Эти устройства представляют собой инфракрасные камеры с математическими алгоритмами, которые позволяют откалибровать взгляд человека, то есть угадать точную точку, куда смотрит человек, и, на втором этапе, щелкнуть ожиданием (пристальным взглядом), морганием или комбинированием взгляда с внешним переключателем или кнопкой (см. рисунок 6).

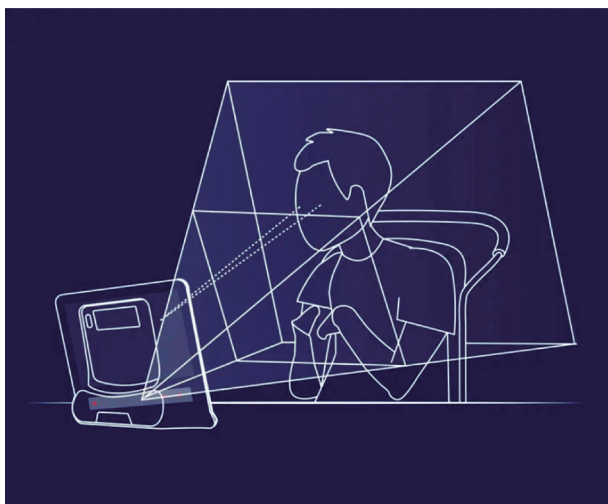


Рис. 6. Иллюстрация компании Tobii Dynavox того, как работают устройства для чтения глаз

Доступ к компьютеру посредством взгляда позволяет установить плавное и полное общение как на формальном, так и на практическом уровне и является единственным эффективным вариантом для людей с очень ограниченной подвижностью. Устройство для чтения глаз отслеживает движения радужной оболочки с помощью двух инфракрасных камер и других датчиков, которые затем преобразуются в сигналы и математические алгоритмы. Он подключается к компьютеру, и его программное обеспечение преобразует движения глаз во входные сигналы, поэтому он работает как обычная мышь и позволяет нажимать на экран. Управление этими технологическими ресурсами подразумевает принятие решений, разработку действий с определёнными целями и специалистов, желающих пройти обучение. Возможности обучения тому, как обращаться с этой технологией взгляда, часто являются даже большим препятствием, чем экономический барьер, поскольку эти устройства обычно не собираются национальными службами общественного здравоохранения. В описанном процессе активно участвует образовательное сообщество: сами дети, а также родители, логопеды и даже частые собеседники, такие как другие родственники и опекуны.

Технология отслеживания движения глаз приписывается шведской компании, расположенной в Гетеборге в 1999 году, а устройства для чтения

глаз впервые были использованы в здравоохранении в Швеции в 2007 году, по данным компании Tobii, и одним из самых больших препятствий на пути к тому, какие преимущества получают пользователи от этих систем, являются их высокие экономические затраты, хотя с годами они снижаются. В 2019 году устройства для чтения глаз были включены в портфель общих услуг Министерства здравоохранения Испании. В частности, предоставление этих устройств в соответствии с законом предусматривается для «пациентов с тяжёлыми нейромоторными расстройствами, с тяжёлым поражением обеих верхних конечностей и невозможностью устного или письменного общения, в основном для пациентов с боковым амиотрофическим склерозом, тромбозом основной артерии, детский церебральный паралич, черепно-мозговая травма и мостовой миелолиз» (Приказ SCB/480/2019 от 26 апреля).

Уже обнародованы данные об использовании описанной технологии офтальмологами для улучшения качества жизни пациентов. В работе Londral и др. (2015) показано положительное влияние на общение не только у человека, который его использует, но и на его окружение, на тех, кто за ним ухаживает. Другие исследования, такие как Ball и др. (2004) и Калигари и др.. (2015) утверждают, что позитивные изменения качества жизни, а также автономия и участие обусловлены рекомендацией специалистов по внедрению SAAC в жизнь.

В Испании продаются устройства для чтения глаз только двух марок: Tobii Dynavox и испанская марка Irisbond. Хотя есть и другие, во всём мире зарегистрировано не более десяти брендов устройств для чтения глаз в области здравоохранения. Некоторые компании, такие как Microsoft, протестировали доступ с помощью взгляда через веб-камеру, хотя эти приложения не обладают необходимой точностью, которая требуется людям с ограниченными физическими возможностями, поскольку устройства вычисляют фиксацию глаз с учётом аномальных движений головы, косоглазия, различных условий освещения, зрачка, цвет и т.д.

Ридеры до недавнего времени были привязаны к среде Windows. Между тем, в 2019 году небольшая немецкая компания EyeV запускает Skyle как первый айтрекер для iOS, распространяемый с 2020 года компанией Inclusive Technology из Великобритании. Наконец, в конце 2021 года Tobii запускает TDPilot в качестве альтернативы iPad. Несмотря на новаторство, которое они представляют, у нас до сих пор нет отзывов профессионалов или пользователей об этих двух устройствах. Для чтения глаз требуется оборудование средней или высокой производительности из-за большого объёма данных, которые они обрабатывают. Это связано с количеством изображений, используемых в коммуникационном программном обеспечении, с которым используется устройство взгляда. В новых поколениях считыватели глаз способны сами обрабатывать данные, а не передавать их компьютеру. С обновлением iOS13 и улучшениями специальных возможностей Assistive Touch доступ к этой операционной системе был разрешён через устройство отслеживания взгляда.

Последние технологические достижения в области чтения глаз для общения также связаны с условиями освещения, в которых находится человек, который их использует. В первом поколении считывателей инфракрасное излучение не позволяло использовать их при дневном освещении или на солнце. В новом устройстве Tobii (устройства i-Series и устройство PCEye) и в новом Smartbox (Lumin-i), разработанном компанией Smart Eye, это свойство отмечается как преимущество по сравнению с предыдущими поколениями. Уже более 20 лет компания Smart Eye

разрабатывает искусственный интеллект (ИИ) в виде технологии отслеживания взгляда, которая понимает, поддерживает и предсказывает намерения и действия человека, тщательно изучая движения глаз, лица и головы. О новом Lumin-i, запущенном в июле 2021 года, компания заявляет: «Lumin-i требуется всего 25 мс (миллисекунд), чтобы отреагировать на ваш взгляд, снимая 60 кадров в секунду со временем восстановления взгляда всего 17 мс. Lumin-i также имеет механизм (большую гусеницу размером 35 X 30 X 65 см.), которая позволяет много двигаться головой, предоставляя пользователям с ограниченной мобильностью надёжное и мощное решение для общения и независимости».

### 3. ИНТЕРФЕЙС МОЗГ-КОМПЬЮТЕР (ВСИ) И ИХ ВЗАИМОСВЯЗЬ

BCI — это системы, готовые улавливать, интерпретировать и обрабатывать то, что думает субъект и каковы могут быть его намерения, благодаря импульсам, которые производятся в их мозгу. Они позволяют создать коммуникационный мост между нашим мозгом и ближайшим окружением без использования обычных механизмов периферических нервов или мышц. Существуют интерфейсы мозг-компьютер (ИМК) и интерфейсы мозг-машина (ИМТ). Благодаря подобным системам их пользователи могут произвольно манипулировать активностью своего мозга, генерируя мозговые сигналы, без использования движений.

Как показывает работа Хосе Андреса Гонсалеса (см. его выступление в проекте «Способность к общению») с пациентами с боковым амиотрофическим склерозом (БАС) и раком горла или ларингэктомии была проведена большая работа по разработке соответствующих речевых технологий.

Большинство работ посвящено автоматическому распознаванию речи и проблемам, возникающим при распознавании речевых сигналов людей с дизартрией или своеобразными голосами, были разработаны модификации программ, чтобы они могли распознавать голоса подобных людей. С другой стороны, выделяются последние достижения в области «молчаливых устных интерфейсов» (González et al. 2016). Это набор технологий, которые пытаются восстановить речь людей, потерявших голос в результате нейродегенеративного заболевания или несчастного случая. С помощью алгоритмов ИИ сигналы, генерируемые человеческим телом, опознаются, пока эти люди «говорят молча».

Группа биомедицинской инженерии Университета Валенсии работала над проектированием, разработкой и оценкой систем BCI как в области обработки сигналов, так и в области программирования вспомогательных приложений для людей с ограниченными физическими возможностями (Martínez-Cagigal et al. 2019). В нескольких проектах сигнал электроэнцефалограммы использовался для пациентов с церебральным параличом посредством вызванных зрительных сигналов. Система BCI специально ориентирована на систему дополнительной связи на основе пиктограмм. Это означает, что пользователи во время ЭЭГ

фокусируются на точке на экране, которая мигает с постоянной частотой, а ВСИ наблюдает за синусоидой на ЭЭГ с этой частотой, чтобы определить конкретную точку, на которую смотрит пациент. Пользователь получает неожиданный визуальный стимул от случайно мигающих огней, что позволяет генерировать реакции (Martínez-Cagigal, 2020). Пользователи, изученные Мартинес-Кагигал, из разных групп с ограниченными двигательными возможностями, контролировали устройства, работали в Интернете и писали слова с помощью этого метода, что свидетельствует о значительном прогрессе, поскольку подобные тесты ранее проводились только с людьми без инвалидности.

Другие американские компании, такие как Synchron или Neuralink, также работают над ВСИ и возможностям расширения коммуникации. Устройство Synchron Stentrode обеспечивает компьютерную связь благодаря интеграции небольшого электродного устройства, предназначенного для имплантации рядом с моторной корой. Вмешательство проводится под общей анестезией с помощью катетера, направляемого по 3D-изображениям, через яремную вену благодаря небольшому разрезу на шее. На следующем изображении можно увидеть графическое представление операции.

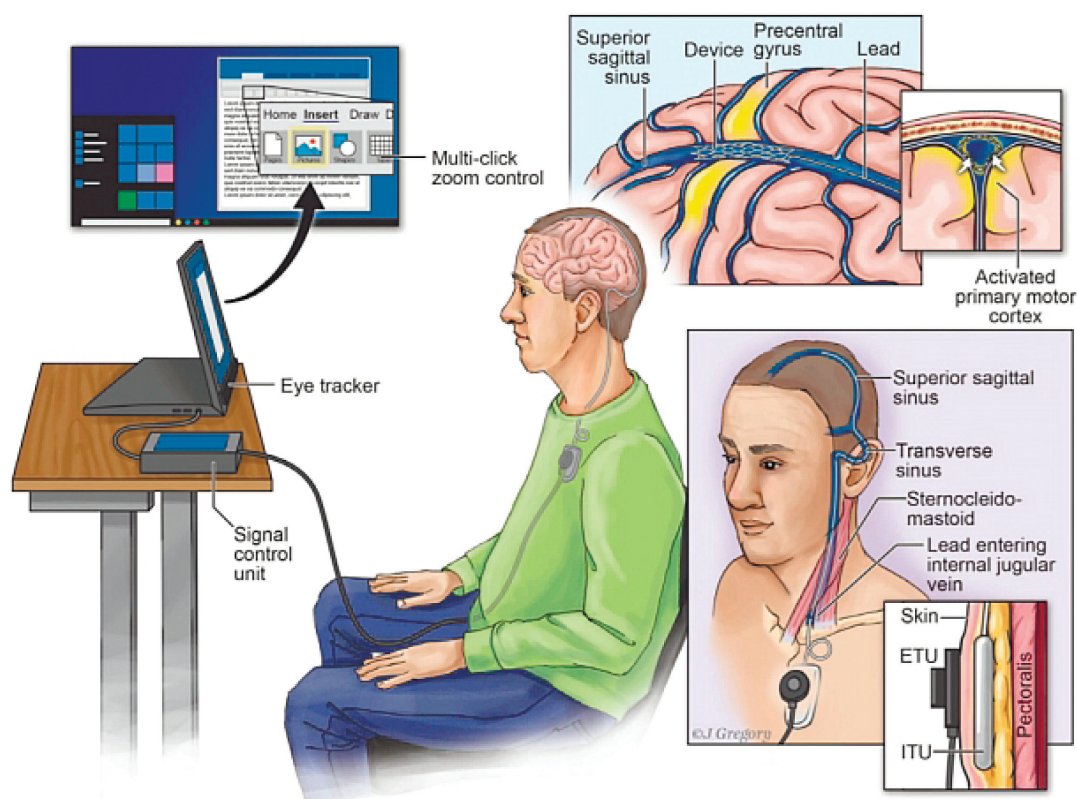


Рис. 7. Работа устройства Synchron Stentrode (Oxley 2020: 103)

Исследования со взрослыми пациентами с БАС, которые тестировали данные устройства, показали, что они могут писать тексты со средней скоростью 20 символов в минуту, поэтому теоретически это может быть путь для будущего развития, более

близкий к тому, что можно было бы придумать в области аугментативной и альтернативной коммуникации.

#### 4. НОВЫЕ МЕТОДОЛОГИИ ВМЕШАТЕЛЬСТВА

Для взрослых с диагнозом нейродегенеративное заболевание, такое как БАС или рассеянный склероз, рак горла, дистония или другие заболевания, влияющие на речь, можно использовать алфавитные доски для общения. За исключением афазии, при этих преимущественно соматических заболеваниях больной может продолжать пользоваться письмом как средством общения. Для перевода письменной системы в устную используются как низкотехнологичные устройства (фреймворки ETRAN), так и виртуальные клавиатуры компьютеров или высокотехнологичные SAAC.

С другой стороны, когда речь идёт о детях с отсутствием речи, мы должны использовать другие коды, такие как фотографии, рисунки или пиктограммы. В 1980-х годах была разработана ПКС (Пиктографическая система связи), состоящая из набора пиктограмм, предназначенных для обозначения на подставке для связи с собеседником. Другие библиотеки предложили альтернативы, и мы можем выделить Widgit, Symbolstix, Metacom, ARASAAC или Sclera. Некоторые из этих наборов символов платные, а другие, такие как испанский ARASAAC, находятся в свободном доступе. В некоторых материалах в России используются пиктограммы ARASAAC (см. рис. 3), хотя немаловажным камнем преткновения является культурный факт, поскольку в испанской библиотеке символов отсутствуют многие типичные для русской культуры пиктограммы. На сайте ARASAAC есть несколько русских переводчиков, в том числе Ирина Текоцкая.

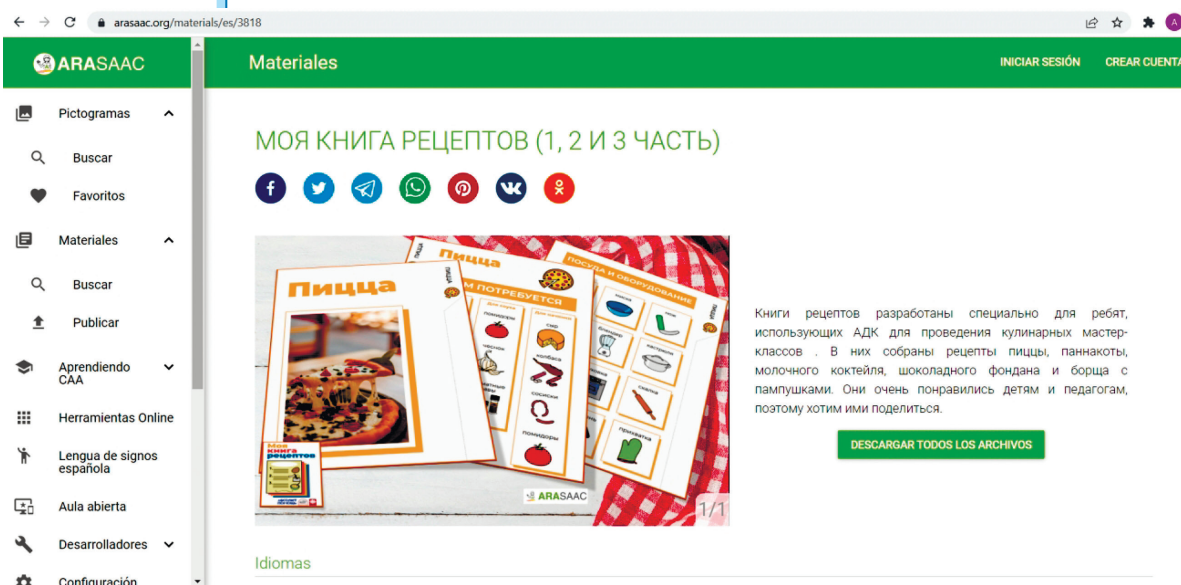


Рис. 8. Материалы на русском языке на сайте ARASAAC

Для большинства детей описанные выше процессы развиваются естественно и спонтанно, если их рост происходит в благоприятной среде и контексте. Типично развивающиеся дети получают постоянную обратную связь о форме, значении и использовании своих вербальных произведений на протяжении всего периода раннего развития. Эта обратная связь позволяет им восстанавливать и улучшать свои первые языковые выбросы. Во время разговора взрослые приспосабливают свою речь к детскому лингвистическому уровню, используя более короткие предложения, паузы и повторения (Clark, 2014). Эти поддерживающие стратегии, такие как модели, переформулировки и помощь опытных собеседников, оказывают решающее влияние на развитие различных языковых навыков ребёнка (Clark, 2018; Savaldi-Harrusi & Soto, 2018). Таким образом, на языковое развитие напрямую влияет качество языковой среды (Weisleder & Fernald, 2013).

Хотя типично развивающиеся дети могут полагаться на речь для участия в разговорах и взаимодействиях, это часто не помогает людям со сложными коммуникативными потребностями. Дети с отсутствием речи или с неразборчивой речью, или дети, которые могут произносить только несколько слов, должны полагаться на такие стратегии, как вспомогательная коммуникация, то есть использование внешних средств связи, таких как динамические коммуникаторы с голосовым выводом или коммуникационные панели, среди прочего, чтобы «помочь» совместному конструированию смысла во время взаимодействия со своими собеседниками (Smith & Murray, 2016). Вспомогательное общение позволяет развивать язык альтернативными способами естественной речи, если ребенок подвергается разнообразному языковому опыту, что является ключом к языковому развитию (Tomasello, 2003; von Tetzchner & Stadskleiv, 2016). Дети, использующие вспомогательную коммуникацию, часто испытывают несоответствие между модальностью ввода языка — разговорным языком, с которым они сталкиваются, — и их собственным лингвистическим опытом (Smith & Grove, 2003). Кроме того, они часто имеют ограниченный доступ к профессионалам, которые показывают им, как должна использоваться их система связи.

Таким образом, в отличие от детей, которые используют естественную речь, дети, которые развивают язык с помощью вспомогательной коммуникации, зависят от сознательных и преднамеренных усилий родителей, учителей и других специалистов. Несколько исследований показали, что обучение собеседников так же важно, как и обучение самих пользователей. Обучение собеседников (родителей, учителей и одноклассников) стратегиям моделирования и общения (Kent-Walsh et al., 2015) необходимо для развития речи при обучении говорить с помощью SAAC с помощью технологий. Эти дети нуждаются в доступной им языковой стимуляции с помощью техники или стратегии моделирования, в которых собеседники указывают на символы (обычно пиктограммы) во время разговора с ребёнком, нуждающимся в этой поддержке. Моделирование, также называемое дополнением лингвистического ввода, было описано как эффективный механизм поддержки понимания и выразительной речи (O'Neill et al., 2018).

## 5. ВЫВОДЫ

Дополняющие и альтернативные коммуникационные системы предоставляют людям с нарушениями речи доступ к общению, соучастию и отдыху. С помощью специального программного обеспечения они могут развивать невербальный язык



детей, независимо от того, существуют ли у них физические или ментальные недостатки. В первом случае, когда руки не могут использовать переключатели или джойстики, считыватели глаз представляют способ доступа, который радикально меняет жизнь таких людей и их окружения, давая им лингвистический код, сначала пиктографический, а затем буквенный для общения. Считыватели для глаз также помогают осуществлять доступ к компьютеру с помощью функции мыши для глаз («мышь для глаз»), чтобы дети с ограниченными возможностями могли получить доступ к образованию, социальным сетям и другим программам для отдыха или работы. Поскольку сложно найти взрослых людей, которые могут рассказать о своём опыте использования подобных технологий, и в проекте *Sarases de Comunicag* мы взяли интервью у двух пользователей, которые учились и работают благодаря им. В представленной статье мы дали несколько сжатых описаний новейших функций, которые позволяют более адекватно и корректно реализовать подобные системы применительно к программному обеспечению и средствам чтения для глаз. Мы также рассмотрели новейшие методологии вспомогательного естественного языка, то есть то, как использовать это программное обеспечение, чтобы развитие языка и производство детской речи шло адекватно и система не была заброшена. Наконец, хотя система Интерфейс мозг-компьютер еще не является частью технологий дополнительных коммуникаций, всё говорит о том, что через одно-два десятилетия они могут стать ещё одной альтернативой доступа и открыть путь для людей с ограниченными возможностями, которые не могут пользоваться зрением, а также читателей с проблемами со зрением.

### Литература

1. *Beukelman, D. & Light, J.* (2020). *Augmentative and alternative communication: Supporting children and adults with complex communication needs* (5th Edition), Paul H. Brookes Publishers. ISBN 9781598573749
2. *Митрофанова, И.И.* (2019). Клиповое мышление: реальность и перспективы. *Речевые технологии 2019*.
3. *Чучупал, В.Я.* (2020). Нейросетевые модели языка для систем распознавания речи. *Речевые технологи, 1-2-2020*. (información para ti: va de sistemas neuronales de reconocimiento de lenguaje hablado, también para intérpretes automáticos)
4. *Ляксо Е.Е., Фролова О.В.* (2020). Акустические маркеры аутизма. *Речевые технологи, 1-2-2020*. (marcadores acústicos para el diagnóstico en función del nivel del habla y tratamiento del autismo)
5. *Вашкевич М.И., Рушкевич Ю.Н.* (2020). Обзор систем автоматического детектирования речевых нарушений пациентов с боковым амиотрофическим склерозом (БАС). *Речевые технологи, 1-2-2020*. (detección automática de trastornos del habla en pacientes con esclerosis amiotrófica lateral)
6. Graves, A. Mohamed y G. Hinton, «Speech recognition with deep recurrent neural networks», de *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013.
7. *Taylor, P.* (2009).- *Text-To-Speech synthesis*, Cambridge, Cambridge University Press.

8. *Mirenda P.* (2013) Augmentative and Alternative Communication. In: Volkmar F.R. (eds) Encyclopedia of Autism Spectrum Disorders. Springer, New York, NY. [https://doi.org/10.1007/978-1-4419-1698-3\\_486](https://doi.org/10.1007/978-1-4419-1698-3_486)
9. *Schlosser, R. W.* (2003b). The Efficacy of Augmentative and Alternative Communication: Towards Evidence-Based Practice. Baltimore: Paul Brookes.
10. *Schlosser, R. W., & Blischak, D. M.* (2001). Is there a role for speech output in interventions for persons with autism. Focus on Autism and Other Developmental Disabilities, 16, 170–176.
11. *Рыскина В.* (ред.-сост.) Альтернативная и дополнительная коммуникация в работе с детьми и взрослыми, имеющими интеллектуальные и двигательные нарушения, расстройства аутистического спектра / СПб.: Издательско-Торговый Дом «Скифия», 2016. — 288 с. — ISBN 978-5-00025-076-1.
12. *Martínez-Cagigal, V.; Santamaría-Vázquez, E.; Gomez-Pilar, J.; Hornero, R.* A brain-computer interface web browser for multiple sclerosis patients. En: Ayman El-Baz, A.; Suri, Jasjit S. Neurological Disorders and Imaging Physics, Volume 2: Engineering and clinical perspectives of multiple sclerosis. IOP Science; 2019. p. 1-31
13. *Martínez-Cagigal, V.* (2020). Toward practical p300-based brain-computer interfaces: asynchrony, channel selection and assistive applications. Tesis doctoral dirigida por Roberto Hornero Sánchez (dir. tes.). Universidad de Valladolid (2020).
14. Orden SCB/480/2019, de 26 de abril, por la que se modifican los anexos I, III y VI del Real Decreto 1030/2006, de 15 de septiembre, por el que se establece la cartera de servicios comunes del Sistema Nacional de Salud y el procedimiento para su actualización. Boletín Oficial del Estado, 101, de 27 de abril de 2019, páginas 43018 a 43028. Tomasello, M. (2003). Constructing a language. A Usage-Based Theory of Language Acquisition. Harvard University Press.
15. *Von Tetzchner, S. y Stadskleiv, K.* (2016). Constructing a language in alternative forms. En: M Smith y J. Murray (Eds.). The Silent Partner? Language, Interaction and Aided Communication (pp. 17-34). J & R Press Ltd.
16. *Tomasello, M.* (2003). Constructing a language. A Usage-Based Theory of Language Acquisition. Harvard University Press.
17. *Weisleder, A., y Fernald, A.* (2013). Talking to Children Matters: Early Language Experience Strengthens Processing and Builds Vocabulary. Psychological Science, 24(11), 2143–2152.
18. *Oxley TJ, et al.* (2020). J Motor neuroprosthesis implanted with neurointerventional surgery improves capacity for activities of daily living tasks in severe paralysis: first in-human experience NeuroIntervent Surg 2021;13:102–108. doi:10.1136/neurintsurg-2020-016862
19. *Smith, M. y Murray, J.* (2016). Introduction. En: M. M. Smith y J. Murray (Eds.). The Silent Partner?. Language, Interaction and Aided Communication (pp. 1-16). J & R Press Ltd.
20. *Savaldi-Harrusi, G. y Soto, G.* (2018). Changes in adult scaffolding in conversations with children who uses aided AAC. Communication Matters, 32(3), 4-6.
21. *Clark, E. V.* (2018). Conversation and Language Acquisition: A Pragmatic Approach. Language Learning and Development, 14(3), 170–185.
22. *Clark, E. V.* (2014). Pragmatics in acquisition. Journal of Child language, 41(S1), 105-116.
23. *Fernández-García, S., Gabriela Dumitrache, C. y González-López, J.A.* (2021). Acoustic analysis of the voice in patients with Parkinson's disease and hypokinetic dysarthria, Revista de Logopedia, Foniatría y Audiología, Volume 41, Issue 3, 2021, Pages 142-150
24. *Gonzalez J.A., Cheah LA, Gilbert JM, et al.* (2016) A silent speech system based on permanent magnet articuloigraphy and direct synthesis. Comput Speech Lang 39: 67-87. doi: 10.1016/j.csl.2016.02.002
25. *O'Neill, T., Light, J., & Pope, L.* (2018). Effects of interventions that include aided augmentative and alternative communication input on the communication of individuals with complex communication needs: A meta-analysis. Journal of Speech, Language, and Hearing Research, 61, 1743–1765. doi:10.1044/2018\_JSLHR-L-17-0132



26. Kent-Walsh, J., Murza, K. A., Malani, M. D., & Binger, C. (2015). Effects of Communication Partner Instruction on the Communication of Individuals using AAC: A Meta-Analysis. *Augmentative and Alternative Communication*, 31(4), 271–284.
27. Weisleder, A., y Fernald, A. (2013). Talking to Children Matters: Early Language Experience Strengthens Processing and Builds Vocabulary. *Psychological Science*, 24(11), 2143–2152.

## TECNOLOGÍAS DEL HABLA PARA EL DESARROLLO DE LA COMUNICACIÓN EN NIÑOS CON SIN LENGUAJE ORAL

*Ana Medina Reguera, PhD, University Pablo de Olavide, Department Translation & Interpreting, Ctra. de Utrera, 1, 41013 Sevilla - Spain, anamedina@upo.es*

In this paper we present some speech and communication technologies that are changing the quality of life of nonverbal children, such as children with cerebral palsy and related conditions. In particular, we point out the technological advances related to augmentative and alternative communication (AAC) software, which allow a better linguistic configuration of pictographic symbols systems. In addition, there have been advances in the access methods for these systems, such as eye-trackers and Brain-Computer-Interfaces. We tackle the technological advances related to augmentative and alternative communication (AAC) software, which allow a better linguistic configuration of pictographic systems. In addition, there have been advances in the methods of accessing these systems, such as eye-readers and Brain-Computer-Interfaces. From both groups we will address aspects of artificial intelligence which are responsible for a direct improvement of the expressive ability of AAC users: on the one hand, symbol combination and linguistic prediction in virtual keyboards; on the other hand, artificial intelligence that allows improving mathematical models for gaze access to the computer. We will rely, in addition to scientific studies on the subject, on our scientific transfer project 'Capaces de Comunicar' [Able to communicate], a project funded by the Spanish Foundation for Science and Technology of the Spanish Ministry of Science and Innovation.

*• augmentative and alternative systems • speechless children • artificial intelligence technologies*

En este trabajo presentamos algunas tecnologías del habla y de la comunicación que están cambiando la calidad de vida de los niños sin lenguaje oral, como niños con parálisis cerebral y enfermedades afines. En concreto, abordamos los avances tecnológicos relativos a los software de los sistemas aumentativos y alternativos de la comunicación (SAAC), que permiten una mejor configuración lingüística de los sistemas pictográficos. Además, se han producido avances en los modos de acceso a estos

систemas, tales como los lectores oculares y los Brain-Computer-Interfaces. De ambos grupos señalaremos aspectos de la inteligencia artificial responsables de una mejora directa de la capacidad expresiva de los usuarios de SAAC: por un lado, la combinación de pictogramas y la predicción lingüística en teclados virtuales; por otro, la inteligencia artificial que permite mejorar los modelos matemáticos para el acceso por mirada al ordenador. Nos apoyaremos, además de en los estudios científicos sobre el tema, en nuestro proyecto de transferencia científica 'Capaces de Comunicar', un proyecto financiado por la Fundación Española de la Ciencia y la Tecnología del Ministerio de Ciencia e Innovación de España.



# Понятие «устойчивый туризм» в социальной, экономической, экологической и лингвистической сферах

*Изабель Гальего Гальярдо, кафедра французской и английской филологии, исследовательская группа HUM-485, преподавание иностранных языков: материалы для разработки новой учебной программы, Кадисский университет, Испания*

Статья посвящена вопросам изучения устойчивого туризма с помощью речевых технологий, а также необходимости введения данного понятия во все сферы и структуры туристического сектора. Представленная тема сопряжена с определёнными трудностями при изучении и применении на практике. В статье предлагается объяснение различий между определениями «туризм» и «устойчивый туризм», описание их исторического развития, обсуждается вопрос о положительном и негативном воздействии туризма. Учитываются последствия туризма как на социальном, так и на экономическом уровнях, чтобы проанализировать воздействие на природу и природоохранную деятельность. Анализ туристического сектора показывает трудности, связанные с реализацией практических мер для сохранения окружающей среды, которые интегрируют устойчивость как культуру туристических проектов. Анализ эколингвистической тематики, формирование лексического корпуса на немецком языке и развитие экологической символики в первую очередь призваны подчеркнуть необходимость превращения темы в приоритетную на дискуссионных форумах. Можно сформулировать предложения по устойчивому развитию для надлежащего управления и включения его в туристическую практику. Лингвистические исследования и речевые технологии помогают добиться позитивной осведомленности населения, включая туристических агентов, которые обычно не задаются вопросом, что произойдет, когда мир однажды перестанет одаривать нас тем, к чему мы так привыкли и получали все это время безвозмездно.

• *эколингвистика* • *туризм* • *устойчивость* • *охрана окружающей среды*  
• *экология*.

## WAS IST TOURISMUS?

Tourismus wird als Reise in einen anderen, mehr oder weniger fremden Ort definiert, die mehr als 24 Stunden andauert, jedoch zeitlich begrenzt ist,

und wieder an den Ausgangsort zurückführt. Eine Reise kann aus geschäftlichen, gesundheitlichen oder Studienzwecken unternommen werden. Reisen werden von Menschen aus allen Gesellschaftsgruppen unternommen. Sie finden in größerem bis massenhaftem Ausmaß, mittels bestimmter Dienstleistungen statt. Der Tourismussektor ist durch ein schier grenzenloses Wachstum gekennzeichnet und stellt inzwischen einen der weltweit bedeutendsten Wirtschaftsfaktoren dar. Alljährlich finden Wanderbewegungen von Millionen von Menschen statt, die fast alle Länder der Welt erreichen. An dieser Stelle soll auf das Phänomen des *Massentourismus* hingewiesen werden, das im weiteren Verlauf des Artikels noch einmal aufgegriffen wird. Der internationale Tourismus stellt weltweit 8,3% aller direkten und indirekten Arbeitsplätze und erwirtschaftet rund 11% des globalen BIP (vgl. Hopfinger 2007: 724). Im Jahr 2000 reisten weltweit insgesamt 698,3 Millionen Menschen in andere Länder. In den kommenden Jahren ist mit einem weiteren Anstieg der Anzahl der Reisenden (Prognose für das Jahr 2020: 1,65 Mrd.) und den dafür aufgebrauchten Kosten zu rechnen.

Laut der weltweiten Tourismusorganisation (UNWTO), im 2019, fast zwei Jahrzehnte später, belief sich die Zahl der internationalen Ankünfte von Touristen der Welt auf 1500 Millionen Menschen, was einen Anstieg von 4% gegenüber 2018 entspricht. Das bedeutet, dass der Tourismussektor 1 von 10 Jobs in der Welt anbietet und 7% der globalen Exporte und mehr als 10% des globalen BIP ausmacht. Ebenso wurden Wachstumswartungen für das Jahr 2020 zwischen 3% und 4% geschätzt. Das Phänomen der COVID 19 und die Maßnahmen, die jeder Staat ergriffen hat, um die Verbreitung des Virus zu verhindern, haben diese Prognosen geändert, und zum negativen Wachstum zwischen 1% und 3% im Jahr 2020 für internationale Besuche geführt. Es ist ein beispielloser Fall, vergleichbar mit der globalen Wirtschaftskrise 2009, da sich damals internationale Touristenankünfte um 4% reduziert haben, auch vergleichbar mit dem SARS-Ausbruch im 2003, der nur ein Rückgang von 0,4% ergab.

Tourismus, Fremdenverkehr und Reisen sind im Laufe der Zeit zu einem festen Bestandteil des Lebensstils und der Lebensbedingungen in den Industrienationen geworden. Der Stellenwert des Tourismus und der damit verbundenen Freizeit hat in der heutigen Gesellschaft stark zugenommen. Es werden beträchtliche Summen „für die schönste Zeit des Jahres“ (Hopfinger 2007: 728) ausgegeben. Dies ist u.a. durch steigendes Einkommen, zunehmende Freizeit, abnehmende Arbeitszeit und bessere Fremdsprachenkenntnisse möglich geworden. Wobei zu beachten ist, wie lange sich dieser Trend, in nun mehr kritischen Zeiten, fortsetzen kann.

## AUSWIRKUNGEN UND FOLGEN DES TOURISMUS

Tourismus stellt einerseits einen äußerst wichtigen Wirtschaftsfaktor dar, andererseits bringt er jedoch auch zahlreiche negative - aber auch positive - Auswirkungen und tiefgreifende Veränderungen mit sich. Denn „[...] sowohl räumlich als auch sozial betrachtet profitieren nicht alle Regionen der Welt und nicht alle Bevölkerungsschichten in gleicher Weise vom Tourismus [...]“ (Hopfinger 2007: 713) Auf der einen Seite stehen Anstöße zur Verbesserung der Lebensbedingungen, auf der anderen Seite jedoch gravierende Probleme. Diese Konsequenzen betreffen sowohl den sozialen und ökonomischen als auch den ökologischen Bereich. Exemplarisch seien hierfür soziokulturelle Konfrontationen zwischen Reisenden und Einheimischen, Veränderungen von sozialen Strukturen und deren Beeinträchtigung, bis hin zu Zerstörung, von Flora und Fauna genannt. Zudem erlebt der internationale Tourismus in Entwicklungsländern eine



fortschreitende Diversifizierung, beispielsweise Wüsten- und Safarireisen, Aufenthalte in Schönheitsfarmen oder Golfen und Mountainbiking. Es ist jedoch fraglich, inwieweit diese neuartigen Angebots- und Nachfrageformen einen wirtschaftlichen Impuls für die Zielregion darstellen und ob sie kulturell und sozial vereinbar sind. (vgl. Hamza 2000: 4)

Zwischen Tourismus, Kultur, Wirtschaft und Natur bestehen erhebliche Zusammenhänge. Werden Natur und Kultur zerstört, schwindet die Attraktivität für den Tourismus. Daher muss Tourismuspolitik zu einem gewichtigen Teil gleichzeitig auch Umwelt- und Kulturpolitik sein. Auf die Auswirkungen des Tourismus der verschiedenen Bereiche soll im Folgenden explizit eingegangen werden.

### **SOZIALE EBENE**

Im sozialen Bereich sind touristische Begleiterscheinungen, wie Wanderungsbewegungen, Zerstören von Familienbindungen und Sozialgefügen, Auftreten von Konkurrenz und zahlreiche Einflüsse auf die traditionelle Struktur und Kultur zu nennen.

Als erstes soll auf die Wanderungsbewegungen eingegangen werden. Die Tourismusbranche stellt einen Großteil der Arbeitsplätze zur Verfügung, dadurch kommt es zu Abwanderungen in die neuen, touristischen Brennpunkte. Wanderungsbewegungen machen dabei keineswegs an Landesgrenzen halt. Diese Bewegungen haben teils Auswanderungen ganzer Familien zur Folge, häufig aber nur einzelner, jüngerer Familienmitglieder. Dies ruft zunächst eine räumliche, dann jedoch auch eine reale Lösung der Familienbindung hervor. Darüber hinaus kommt es zu massiven Land-Stadt-Wanderungen, da sich viele Betroffene ein (besseres) Einkommen aus dem Fremdenverkehr in der Stadt erhoffen. Nicht selten wird die ortsansässige Bevölkerung sogar auch zwangsumgesiedelt. Ein weiterer Konfliktpunkt innerhalb einer Familie kann auch das Erschüttern der traditionellen Struktur bzw. der Familienhierarchie sein. Jüngere Familienmitglieder verdienen beispielsweise häufig mehr als das Familienoberhaupt oder besetzen eine höhere Position im Arbeitsleben. (vgl. Vorlaufer 2003: 10) Vorausgesetzt, dass leitende Positionen überhaupt an Einheimische vergeben werden. Denn meist werden diese von einem Personenkreis besetzt, der eine entsprechende Ausbildung (im Ausland) und die notwendigen Sprachkenntnisse vorweisen kann, d.h. die Spitzenpositionen sind von ausländischen Führungskräften besetzt. Hier wird von *wirtschaftlichem Kolonialismus* gesprochen. (vgl. Kreisel/Reeh 2006: 17) Des Weiteren verursacht Tourismus, vor allem, wenn er einzige oder wichtigste Einnahmequelle einer Region ist, Konkurrenzgefühle innerhalb der einheimischen Bevölkerung. Die verschiedenen Anbieter buhlen um den Kunden Tourist, um ihn bestenfalls für die eigenen Angebote zu gewinnen. (vgl. Hamza 2000: 9)

Die Einflüsse auf Kultur und Tradition sind zahlreich. Häufig wird behauptet, der Tourismus hätte eine bewahrende Funktion auf die einheimische Tradition, doch es muss beachtet werden, dass Brauchtum und traditionelle Kultur

teilweise kommerzialisiert werden. Das Brauchtum, durch den Tourismus reaktiviert wird, hat oft nichts mehr mit den ursprünglichen Traditionen zu tun. Brauchtum, sei es im Bereich Kunsthandwerk, Musik oder Tanz, ist darüber hinaus meist nur noch dort erhalten, wo der touristische Einfluss (noch) nicht eingedrungen ist. Die traditionelle Kultur wird zu einem Millionen-Geschäft. Es entstehen Souvenir-Supermärkte und "Heilige Stätten", wie zum Beispiel Buddha Tempel, die sogar für die Öffentlichkeit geöffnet werden, um eine zusätzliche Einkommensquelle durch Eintritt und Führungen zu erlangen. Nicht selten führt der Massentourismus sogar zu massiven Menschenrechtsverletzungen, wie zum Beispiel Kinder- und Frauenprostitution.

Zu Konflikten zwischen den beiden in einer Destination beteiligten Seiten kann es auch dadurch kommen, dass sich die Touristen anders als Einheimische verhalten, sei es in Form ihrer Kleidung oder im Umgang mit der natürlichen Ressource Wasser. Im Fall des Trekking-Tourismus im Hohen Atlas zum Beispiel werden „(...) Frauen, die in Shorts wandern und vielleicht sogar T-shirts mit nackten Schultern zeigen, (...) von der Landbevölkerung als unanständig angesehen“. (Hamza 2000: 8) Die Touristen sind für die Einheimischen fremde Wesen.

## ÖKONOMISCHE EBENE

Im Bereich der Wirtschaft können verschiedene Auswirkungen des Tourismus einander gegenüber gestellt werden. Erstens Einnahmen und Ausgaben. Durch die Touristen gelangen Deviseneinnahmen in das jeweilige Zielland, aber ein Großteil der Einnahmen fließt durch Importe für die Touristen unmittelbar wieder ab oder geht an internationale Firmen, wie Fluggesellschaften, Reiseagenturen, Autovermietungen oder Hotelketten. Weitere Einnahmen können auch durch Steuern und Direkttaxen ins Land fließen. (vgl. Kreisel/Reeh 2006: 16) „In vielen Ländern leistet der Tourismus einen wesentlichen Beitrag zur Verbesserung der Zahlungsbilanz.“ (Vorlaufer 2003: 5) Doch vor allem in den Entwicklungsländern kann auch eine negative Bilanz dadurch entstehen, dass die Ausgaben für den Tourismus schneller als die Einnahmen wachsen. Darüber hinaus kann der Fremdenverkehr eine inflationäre Wirkung mit sich bringen: die touristische Nachfrage treibt die Preise im Zielgebiet in die Höhe, für die Touristen, jedoch auch für die einheimische Bevölkerung.

Als zweites können die Entstehung von Arbeitsplätzen und die Überlastung des Arbeitsmarkts gegenübergestellt werden. Der Tourismus als Arbeitgeber hat, wie erwähnt, einen großen Arbeitskräftebedarf an direkten (Hotelpersonal, Gastronomie) und indirekten (Baugewerbe, Werbeagenturen, Handel) Arbeitsplätzen. Im Jahr 2005 arbeiteten 74,2 Millionen im Kernbereich des Tourismus und weitere 147,3 Millionen im indirekten touristischen Bereich (vgl. Hopfinger 2007: 724). Die Daten vor der Pandemie, im Jahr 2019, zeigten den Wohlstand des Sektors, der 1 von allen vier neuen Arbeitsplätzen garantierte und 334 Millionen Arbeitsplätze anbot - 10,6% des weltweiten Gesamtangebotes auf dem Arbeitsmarkt (vgl. Weltreisen und Tourismusrat, 2020). Doch der lokale Arbeitsmarkt wird kaum entlastet, sondern durch die Sogwirkung, die der Tourismus mit sich bringt, zusätzlich belastet, vor allem dann, wenn sich der Tourismus auf ein städtisches Zentrum konzentriert. Häufig gibt es vor Ort ein Überangebot an unqualifizierten Arbeitskräften. Dies führt folglich zu einer hohen Arbeitslosigkeit. Die Beschäftigungsmöglichkeiten im Tourismussektor sind weiter oft nur saisonal gegeben. Der Fremdenverkehr ist häufig mangels Alternativen eine Art *Monokultur* der jeweiligen Zielregion. Hohe Vulnerabilität und Abhängigkeit entsteht





demzufolge auch, wenn ein Tourismuszentrum massiv vom Zustrom ausländischer Besucher abhängt. Beispiele hierfür sind der Rückgang von Besucherzahlen nach dem 11. September 2001 und der SARS-Epidemie. (vgl. Kreisel/Reeh 2006: 17). In diesem Sinne haben sich der Schaden, der durch Kriegskonflikte und terroristische Taten in den Infrastrukturen und der Ausrüstung bestimmter Ziele verursacht hat, die wichtigsten Touristenentscheidungen aus Sicherheitsgründen beeinflusst. Da der syrische Bürgerkrieg und die staatliche Zersetzung im Irak der förderliche Kontext waren, so dass Gruppen wie ISIS emblematische Enklaven des hohen historischen Wertes in Städten wie Aleppo, Mosul oder Palmira (Castelo 2020: 116).

Als dritter Punkt sind der (Aus-) Bau von Infrastruktur und die dafür notwendigen Kosten gegeneinander abzuwägen. Der Aufbau oder die Aufbesserung von Infrastruktur dient nicht nur den Reisenden, sondern auch der einheimischen Bevölkerung. Doch der Bau und die Instandhaltung von Flughäfen, Straßen, Schienen, Telefon, Post und Presse ist auch mit erheblichen Kosten verbunden. Wenn einheimische Arbeitskräfte im Bausektor beschäftigt werden, hat der (Aus-) Bau von touristischer Infrastruktur allerdings auch einen Beschäftigungseffekt. (vgl. Kreisel/Reeh 2006: 16)

## ÖKOLOGISCHE EBENE

Dass der Tourismus zahlreiche, schwerwiegende Folgen für die Umwelt hat, ist allseits bekannt. „Die stärksten, in vielen Reiseländern offensichtlichsten Negativeffekte des Tourismus beeinflussen die natürliche Umwelt. Insbesondere in Massentourismus-Destinationen (...)“ (Vorläufer 2003: 11) Während die ökologischen Belastungen durch den Tourismus infolge der geringen Anzahl an Touristen bis Ende der 1950er Jahre vernachlässigbar waren, nahmen die Konsequenzen mit dem einsetzenden und seit Jahren boomenden Massentourismus zu. Die fortschreitende Zunahme des Tourismus und der Trend zu kürzeren, jedoch häufigeren Reisen in immer entferntere Länder führen zu einer verstärkten Belastung der Umwelt. Nicht nur landschaftsbezogener Tourismus und Freizeitsport nehmen oft empfindliche Lebensräume in Anspruch. Es kommt zu ökologischen Belastungen durch neue touristische Infrastruktur, erhöhten Flächen- und Energiebedarf, gesteigerten Ressourcen- und vor allem Wasserverbrauch, zusätzliche Schadstoffbelastung und durch den Anstieg von Abwasser und Müll.

Dabei spielen der Erhalt einer natürlichen Umwelt und einer unberührten Natur eine entscheidende Rolle als Reisemotiv für die Touristen. Laut einer Umfrage ist beispielsweise für mehr als 75% aller Deutschen eine ursprüngliche Naturlandschaft am Urlaubsort sehr wichtig. Aber zu starke Belastungen können dazu führen, dass sich die Touristen vom Zielgebiet abwenden. Tourismus ist folglich einerseits von einer intakten Natur abhängig andererseits beeinträchtigen gerade Freizeit und Tourismus Flora und Fauna. Die negativen Auswirkungen auf die Umwelt lassen sich in drei Teilbereiche unterteilen: Belastungen der Umwelt

durch touristisch motivierten Verkehr, Belastungen durch touristische Infrastruktur (Erschließungsfolgen) und Belastungen durch Nutzungsfolgen, d.h. durch touristische Aktivitäten. Die drei Bereiche werden im Folgenden kurz erläutert.

Mit einem Anteil von ca. 50% an der Verkehrsleistung in Deutschland, dem Land der Reiseweltmeister, stellt der Freizeit- und Urlaubsverkehr den quantitativ bei weitem bedeutendsten Fahrtzweck dar. Auto und Flugzeug sind hierbei die beliebtesten Reisemittel, doch diese sind gleichzeitig auch die größten Umwelt- und Luftverschmutzer (Lärmbelastung, Schadstoffemission, vor allem durch die Freisetzung von Schadstoffen durch den Flugverkehr in der Tropopause, die wesentlich zum Ozonschichtabbau und zum Treibhauseffekt beitragen, usw.). Öffentliche Verkehrsmittel, die besonders umweltfreundlich sind, hingegen verlieren aufgrund von konstanten Preiserhöhungen und dem individuellen Freiheitsaspekt immer mehr an Bedeutung. Hotels, Ferienapartments und -häuser, Campingplätze, Ver- und Entsorgungsinfrastruktur, Parkplätze, Sportboothäfen, Skipisten und Golfplätze sind Beispiele für die ökologische Belastung durch touristische Infrastruktur. Die Landschaft und das Landschaftsbild werden zerstört. Es entstehen sog. Touristenghettos an Küstenabschnitten oder Retortenstädte in den Alpen. Der Skitourismus in den Alpen kann als Exempel für Belastungen durch touristische Aktivitäten angeführt werden. Durch den Bau von immer mehr Skipisten kommt es zu großräumigen Umweltschäden, wie Murenabgänge, Lawinen und Zersiedlung. Ein weiteres Beispiel ist der exzessive Verbrauch der zunehmend knappen Ressource Wasser, wie im Fall der überaus zahlreichen Golfplätze in Andalusien belegt werden kann. Wasserreserven werden weiter vor allem auch durch den Küstentourismus (Duschen, Pools, Bewässern von Parks und Grünanlagen) übernutzt. Belastungen können sowohl bei sanften (Wandern) als auch bei harten Freizeitaktivitäten (Heliskiing) auftreten. Darüber hinaus werden Ruhephasen für die Natur verkürzt oder fallen sogar weg. Eine immer neuere und bessere Qualität von Sportausrüstungen machen Aktivitäten wetter- und jahreszeitenunabhängig.

Abschließend soll auf einige positive Effekte des Tourismus hingewiesen werden. Denn wenn er sinnvoll betrieben wird, kann er sich auch vorteilhaft auf den sozialen, ökonomischen und ökologischen Bereich auswirken. Denn trotz der vielfältigen Belastungen können sich Freizeit und Tourismus positiv auf die Umwelt auswirken. Fremdenverkehr kann dazu beitragen, dass gefährdete Ökosysteme bewahrt und Naturschutzprojekte finanziert werden können. Das Erlebnis intakter Natur kann die Touristen zu mehr Umweltbewusstsein und zu einem schonenderen Umgang mit den natürlichen Lebensgrundlagen motivieren: Naturschutz durch Naturgenuss. Inzwischen ist der Umweltschutz auch von der Tourismuswirtschaft als strategischer Erfolgsfaktor am Markt anerkannt worden. Naturschutz wird zunehmend vermarktet. TUI wirbt beispielsweise mit dem Slogan *Schutz durch Nutzung* und betreibt den Tourismus im Sinne einer Naturschutzorganisation. „Obwohl durch den Tourismus wachsende Umweltbelastungen (mit) verursacht werden, besteht das Paradoxon, dass der Tourismus auch Motor wachsenden Umweltbewusstseins und -schutzes sein kann.“ (Vorlaufer 2003: 11) Wie bereits erwähnt, sind die Schaffung von Einkommen und Beschäftigung, die Integration von Entwicklungsländern in die Weltwirtschaft und die Wiederbelebung einheimischer Lebensstile weitere mögliche Vorteile. Auch die Einnahmen aus der Herstellung von Souvenirs durch den Einsatz einheimischer Materialien und traditioneller Kenntnisse der Bewohner sollen hier nicht unerwähnt bleiben (vgl. Vorlaufer 2003: 7). Häufig bietet der Tourismus den sonst benachteiligten Arbeitskräften Jobs an. Frauen in Entwicklungsländern stellen beispielsweise einen



hohen Anteil an Hotelbeschäftigten. Doch insgesamt gibt es „(...) wenige Gewinner und eine Reihe von Verlierern.“ (Hopfinger 2007: 713)

### WAS BEDEUTET “NACHHALTIGER TOURISMUS“?

Nachhaltiger Tourismus ist umwelt- und sozialverträglicher Tourismus. Alle Tourismusformen und Reiseangebote, die in besonderem Maße zu einem langfristigen wirtschaftlichen Erfolg touristischer Unternehmen und Destinationen, zur Lebensqualität und zur Mitbestimmung der Bevölkerung in den Zielgebieten, zum Erhalt einer intakten und vielfältigen Natur und Kultur und zur Förderung der Lust am Lernen, Entdecken und Genießen beitragen, werden als nachhaltig bezeichnet. Das Konzept des Nachhaltigen Tourismus basiert auf der Definition der Nachhaltigkeit, überträgt sich auf alle Tourismusformen sowie –orte und grenzt sich über die Kriterien der Nachhaltigkeit von nicht nachhaltigen Tourismusformen ab. Soziale, ökonomische und ökologische Ziele müssen gleichzeitig verfolgt werden. Jeder Tourist kann sich beispielsweise vor oder während einer Reise fragen, in welcher Form er Ressourcen benutzt, was seine Devisen bewirken oder welche kulturellen Auswirkungen entstehen. (vgl. Hopfinger 2007: 723) Zu beachten ist jedoch, dass grundsätzlich jede Tourismusform eine Belastung für die Umwelt darstellt. Lediglich die Intensität der Belastung kann durch die Umsetzung der Ziele eines Nachhaltigen Tourismus reduziert werden. „Ökologische Bedenken gegenüber der weltweiten touristischen Expansion bleiben allemal auf der Strecke, weil Konzepte für nachhaltigen Tourismus (...) gegen die Kräfte des Marktes kaum Realisierungschancen besitzen.“ (Hopfinger 2007: 713)

### SYNONYME BEGRIFFE VON NACHHALTIGEM TOURISMUS

Es ist wichtig zu sagen, dass während der Informationsrecherche des vorliegenden Artikels eine spezifische Eigenart im Bereich der Linguistik aufgefallen ist. In der deutschen Sprache existiert ein linguistischer Reichtum an Synonymen bzw. alternativen Begriffen für den Terminus des Nachhaltigen Tourismus. Im Spanischen ist diese Vielzahl an Termini hingegen nicht zu finden. Das semantische (Wort-) Feld des Nachhaltigen Tourismus ist relativ neu und wird darüber hinaus kontinuierlich erweitert und verbessert. Seit den 1960er Jahren, als der Begriff des Ökotourismus entstand, sind weitere Varianten entwickelt worden. Teilweise werden die Synonyme dabei ins Spanische übersetzt, andernfalls lediglich transferiert. Einige der verschiedenen Termini werden im Nachfolgenden erläutert.

“Sanfter Tourismus“, der im Spanischen als *turismo moderado* oder *turismo verde* übersetzt werden kann, ist der Vorläufer des umweltorientierten Tourismus. Der Begriff wurde 1980 vom Zukunftsforscher Robert Jungk geprägt und bezeichnet die Leitvorstellung für eine alternative touristische Entwicklung. Dies bedeutet im Kern, dass freizeitorientierte Umweltgestaltung nur unter der Beachtung der Schonung von Natur,

Landschaft und der ortsansässigen Bevölkerung vollzogen werden sollte. Es werden beispielsweise touristische Touren angeboten, die auf der Idee des sanften Tourismus basieren und mit wenig umweltbelastenden Reisemitteln, wie Fahrrad oder Bus, durchgeführt werden.

Als nächstes Synonym sei der Ökotourismus (span. *turismo ecológico*) genannt. Diese Art des nachhaltigen Tourismus ist eine verantwortungsvolle Form des Reisens in naturnahe Gebiete, die zum Schutz der Umwelt und zum Wohlergehen der lokalen Bevölkerung beiträgt. Es sollen nicht nur negative Auswirkungen auf Natur und Umwelt vermieden, sondern auch aktiv Schutz- und Entwicklungsziele unterstützt werden. Der Terminus "Ökotourismus" ist in den 1960er Jahren in den USA entstanden und soll dem Genuss und der Bewunderung der Natur und den damit verbundenen kulturellen Besonderheiten dienen. Ökotourismus hat weitergehende Anforderungen: Erziehungs- und Bildungsfunktion der Bevölkerung aber auch der Touristen, ökonomische Gewinnoptimierung, Minimierung negativer Auswirkungen, vor allem auf die Umwelt und Förderung des Naturschutzes. Darüber hinaus wird der Ökotourismus von der Tourismusindustrie als Vermarktungsstrategie betrachtet. Das Jahr 2002 wurde beispielsweise als Jahr des Ökotourismus proklamiert. Mit dem Ökotourismus werden Hoffnungen verbunden, ein Konzept zu finden, das spezifisch und stärker als andere Tourismusformen soziale sowie umwelt- und naturschutzorientierte Ansprüche betont. Die Grenzen zum nachhaltigen Tourismus sind dabei fließend, die Hürden zum Massentourismus dagegen sehr hoch.

Des Weiteren muss zwischen Ökotourismus und Naturtourismus unterschieden werden. „Unter Naturtourismus wird eine Art Tourismus verstanden, dessen Hauptzweck das Naturerleben ist.“ (Karrasch 2003: 30) Beim Naturtourismus, der im Spanischen als *turismo natural* bezeichnet wird, steht der Natur- und Landschaftsgenuss im Mittelpunkt.

Auch Abenteuer Tourismus (span. *turismo de aventura*) und Ländlicher Tourismus (span. *turismo rural*) sind synonyme Begriffe für den Nachhaltigen Tourismus. Der Abenteuer Tourismus hat seine Betonung auf sportlich ausgerichtete Aktivitäten, die überwiegend in naturnahen Landschaften ausgeführt werden. Ländlicher Tourismus ist ökologisch und soziokulturell verträglich und erlaubt den Touristen eine Annäherung an die naturräumlichen Gegebenheiten der jeweiligen Destination. Dabei werden sowohl Landwirtschaft als auch Naturerfahrung, Regionalkultur und das soziale Gefüge unterstützt. Ein Beispiel dieser nachhaltig betriebenen Tourismusform ist der Urlaub auf dem Bauernhof. Abschließend seien weitere Synonyme genannt: Agrotourismus, alternativer Tourismus, Biotourismus, Kulturtourismus, Wellness Tourismus und Wildnistourismus.

## LÖSUNGSANSÄTZE DURCH MITARBEIT

Damit die genannten Tourismusformen in der Realität des heutigen Tourismus ausgeführt werden können, müssen Touristen, aber auch Einwohner zunächst für ein natur- und umweltbewusstes Verhalten sensibilisiert und erzogen werden. Im zweiten Schritt müssen die verschiedenen Konzepte des nachhaltigen Tourismus gewissenhaft umgesetzt werden. In Deutschland gibt es eine Vielzahl von Hilfestellungen für einen bewussten Umgang mit der Umwelt. Den Einwohnern und den Touristen in Deutschland wird die Möglichkeit gegeben, dass jeder einen Beitrag zum Natur- und Umweltschutz leisten kann. Dafür gibt es, neben den bereits erwähnten Termini, einerseits zahlreiche

Symbole und Siegel, wie zum Beispiel der *Blaue Engel* oder verschiedene Pfandsymbole. In Deutschland ist Mülltrennung obligatorisch. Auch gibt es Umweltplaketten, um die Luftverschmutzung in (Innen-) Städten zu minimieren und die Schulferien werden aus Umweltgründen in jedem Bundesland anders gelegt. Nachfolgend werden einige Symbole erläutert:



Der *Grüne Punkt* ist inzwischen ein europaweites Kennzeichnungssystem (von Einwegverpackungen) im Verpackungsrecycling. Das Symbol wurde 1991 in Deutschland entwickelt und ist seit 1997 eine EU-Richtlinie.



Der *Blaue Engel* ist ein Symbol für besonders umweltschonende Produkte und Dienstleistungen. Es wurde 1978 in Deutschland ins Leben gerufen und macht beispielsweise Papier kenntlich, das zur Verwendung von Recyclingpapier (Altpapier) verwendet wird.



Als Mehrwegpfand werden Abgaben auf Flaschen und Gefäße bezeichnet, die wiederverwendet werden können. Im Gegensatz zum Einwegpfand, bei dem Verpackungen und Flaschen zerkleinert und weiterverarbeitet werden, laufen Mehrwegflaschen nach einer chemischen Reinigung direkt in den Produktionskreislauf zurück. Das Pfand beträgt 8 oder 15 Cent und wird nach (unbeschädigter) Rückgabe zurückerstattet. Bereits seit den 1980er Jahren gibt es in Deutschland ein einheitliches Pfandsystem.



Seit 2006 gibt es in Deutschland des Weiteren auch ein geschlossenes System für Einwegflaschen. Die Abgaben hierfür liegen bei 0,25€. Durch die Verwendung von Geld als Pfand, wird der Verbraucher zur Rückgabe der Plastikflaschen motiviert.



Das Bio Siegel wurde in Deutschland im Jahr 2001 eingeführt. Das Symbol kennzeichnet Produkte aus ökologischem Landbau. Diese dürfen beispielsweise nicht durch und nicht mit gentechnisch veränderten Organismen hergestellt werden und keine Geschmacksverstärker, künstliche Aromen, Farbstoffe und Emulgatoren enthalten. Weiter sind synthetische Pflanzenschutzmittel untersagt und es werden Mindeststall und -freiflächen vorgegeben.



Der Begriff *Fairer Handel* (span. *comercio justo*) wird hauptsächlich bei Waren, die aus Entwicklungsländern in Industrieländer exportiert werden, angewendet. Fairer Handel betrifft landwirtschaftliche Erzeugnisse und Produkte aus traditionellem Handwerk und Industrie. Doch inzwischen wird der Begriff auch im Bereich des Tourismus angewendet und soll ein faires Reisen kennzeichnen.

Das Gütesiegel macht Produkte aus fairem Handel für die Konsumenten erkennbar. Deutschland ist neben Österreich und der Schweiz einer der Initiatoren. Die Fairtrade Labelling Organizations Internacional, eine Dachorganisation für fairen Handel wurde 1997 gegründet und hat ihren Sitz in Bonn, Deutschland.



Ein neuerer Beitrag zur Schonung der Umwelt ist die sog. *Verordnung zum Erlass und zur Änderung von Vorschriften über die Kennzeichnung emissionsarmer Kraftfahrzeuge*. Diese ist am 1. März 2007 in Deutschland in Kraft getreten. Sie ermöglicht in Städten die Einrichtung von Umweltzonen, die die Emission von Feinstaub und Stickstoffdioxid überwachen. Dazu dienen u.a. die bereits erwähnten Umweltplaketten, die PKWs und LKWs in vier Schadstoffgruppen einordnen. Drei dieser Gruppen werden durch Plaketten ersichtlich, die an der Windschutzscheibe angebracht werden müssen. Nur ausgezeichnete Fahrzeuge dürfen in die jeweiligen Umweltzonen, wie Innenstädte, einfahren. In Zonen, in denen eine sehr hohe Feinstaubbelastung festgestellt wird, dürfen nur Fahrzeuge der Schadstoffgruppe 4, mit grüner Plakette, verkehren. Bei einem Verstoß wird ein Bußgeld von 40€ und ein Punkt im Verkehrszentralregister fällig.



## ZUSAMMENFASSUNG UND SCHLUSSFOLGERUNG

„Als Geburtsstunde des modernen Tourismus gilt die erste Pauschalreise, von Thomas Cook 1841 organisiert.“ (Hopfinger 2007: 719) Durch eine erhöhte Massenkauftkraft, die Einführung technischer Neuerungen, wie zum Beispiel Großraumflugzeuge, das Entstehen einer spezifischen Reiseindustrie, die Liberalisierung des grenzüberschreitenden Waren-, Personen- und Dienstleistungsverkehrs und ein verändertes Freizeit- und Reiseverhalten hat sich seit den 1950er Jahren der moderne Massentourismus und gleichzeitig auch der Tourismus in den Entwicklungsländern entwickelt. „Zunächst wurde vor allem mit Bus und Bahn gereist, dann mit dem PKW. Der Flugtourismus entwickelte sich ab 1955 und gewann durch den Charterflugverkehr an Bedeutung.“ (Hopfinger 2007: 720) Die erste Flugpauschalreise war in Deutschland im Jahr 1956 buchbar. Insbesondere in den 1960er Jahren erlebte der Tourismus eine Euphoriephase. Bis in die Mitte dieses Jahrzehnts wurden die Auswirkungen des Tourismus überwiegend als positiv angesehen. Danach beginnt, verursacht u.a. durch die Ölkrise im Jahr 1973 eine Phase der Ernüchterung. Der Tourismus wird zunehmend kritisch hinterfragt. In den 1980er Jahren wird der nötige Handlungsbedarf erkannt und es werden erste politische Regelungen aufgestellt. Im Jahr 1980 werden die Thesen zum „Sanften Tourismus“ von Robert Junkg veröffentlicht. In den späten 1980er und vor allem in den 1990er Jahren explodiert der ohnehin schon boomende Reisemarkt und es entsteht eine Vielzahl von neuen Angebots- und Nachfrageformen. Das Konzept des Nachhaltigen Tourismus und eine nachhaltige regionale Entwicklung stehen im Fokus. Die Reiseintensität erhöht sich auf über 75%. (vgl. Hopfinger 2007: 720)

„Zur Eigenart des Tourismus gehört, (...), Grenzen in Zeit und Raum zu überwinden.“ (Hopfinger 2007: 713) Dies ist ein typisches Phänomen der Globalisierung. Tourismus ist nicht nur ein Produkt oder eine Begleiterscheinung der Globalisierung, sondern auch Motor. Gleichzeitig kann der Fremdenverkehr einerseits die jeweilige Destination stärken, andererseits löst er beträchtliche Veränderungen aus. (vgl. Hopfinger 2007: 713) Die negativen, wie positiven Folgen des Tourismus betreffen sowohl den sozialen und ökonomischen, als auch den ökologischen Bereich. Vor allem die ökologischen Bedenken gegenüber der weltweiten Expansion des (Massen-) Tourismus bleiben allemal auf der Strecke. Mit dem Erstarken der Umweltbewegungen in den 1980er Jahren nahm zwar die Sensibilisierung der Bevölkerung zu, doch das tatsächliche Umweltverhalten hat sich hingegen kaum verändert. Konzepte des Nachhaltigen Tourismus müssen zunehmend beachtet werden. Dafür muss das bestehende Informationsdefizit, zum Beispiel durch Gütesiegel, abgebaut werden. Reiseveranstalter sollten auf umweltfreundliche Reisemittel hinweisen, umweltschonende Freizeitaktivitäten anbieten und in der Destination selber naturbewusst agieren. Beispiele hierfür gehen von einem gebietstypischen Baustil, einem möglichst schonenden Umgang mit Grund und Boden über den Einkauf regionaler Produkte, bis hin zu dem Verzicht auf Portionspackungen beim Frühstück. Des Weiteren sollte jeder Touristen auf umweltschonende Möglichkeiten, wie Handtücher nur bei Bedarf waschen zu lassen, hingewiesen werden.

Nachhaltigkeit, nicht nur im Bereich des Tourismus, betrifft jeden einzelnen unserer Gesellschaft.

### **Bibliographie**

1. *Castello, Vanesa*: „Desafíos y oportunidades para el turismo en el marco de la pandemia COVID 19“, in: CUPEA Cuadernos de política exterior Argentina, 131, 2020, S. 115-118. <https://doi.org/10.35305/cc.vi131.85>
2. *Hamza, Mohamed Ait und Popp, Herbert*: „Trekking-Tourismus im hohen Atlas. Ein Beispiel für nachhaltigen Tourismus in einer Peripherregion?“, in: Geographische Rundschau 52/2, Braunschweig 2000, S. 4-10.
3. *Hopfinger, Hans*: „Geographie der Freizeit und des Tourismus“, in: Gebhardt, Hans u.a. (Hg.): Geographie. Physische Geographie und Humangeographie, München 2007, S. 713-733.
4. *Karrasch, Heinz*: „Galápagos: fragiles Naturparadies und Ökotourismus“, in: Geographische Rundschau 55/3, Braunschweig 2003, S. 26-33.
5. *Kolbe, Carsten*: „Ökotourismus als Chance? Ein Überblick über die Entwicklung des Konzepts und seiner Umsetzung am Beispiel zweier Großschutzgebiete“, in: Tourismus – Report. Symposium zum Internationalen Jahr des Ökotourismus, Eberswalde 2002, S. 10-20.
6. *Kreisel, Werner und Reeh, Tobias*: „Tourismus auf den Inseln des Nördlichen Pazifiks“, in: Geographische Rundschau 58/9, Braunschweig 2006, S. 12-19.
7. *Mundt, Jörn W.*: Tourismus, München/Wien 20063.
8. Organización Mundial del Turismo: Global and Regional Tourism Performance. 2020. <https://www.unwto.org/global-and-regional-tourism-performance> (letzter Zugriff: 1.3.2020)
9. *Spittler, Rolf*: „Nachhaltige Tourismusentwicklung in Großschutzgebieten. Das Verhältnis von Tourismus und Naturschutz“, in: Tourismus – Report. Symposium zum Internationalen Jahr des Ökotourismus, Eberswalde 2002, S. 47-52.
10. *Vorlaufer, Karl*: „Tourismus in Entwicklungsländern. Bedeutung, Auswirkung, Tendenzen“, in: Geographische Rundschau 55/3, Braunschweig 2003, S. 4-13.
11. World Travel and Tourism Council: Economic Impact Reports. 2020. <https://wtcc.org/Research/Economic-Impact> (letzter Zugriff: 7.6.2020)
12. *Wilken, Thomas*: „Der Weg ist das Ziel – Die Europäische Charta für nachhaltigen Tourismus in Schutzgebieten“, in: Tourismus – Report. Symposium zum Internationalen Jahr des Ökotourismus, Eberswalde 2002, S. 65-71.

### **Abbildungsverzeichnis**

*Der grüne Punkt*: <https://www.gruener-punkt.de/de/downloads>  
(letzter Zugriff: 21.01.2022)

*Der blaue Engel*: <https://www.blauer-engel.de/de>  
(letzter Zugriff: 21.01.2022)

*Mehrweg*: <https://www.mehrweg.org/einkaufen/pfand/>  
(letzter Zugriff: 21.01.2022)

*Einweg*: <https://www.ecosia.org/images/?q=einweg%20logo#id=F4A8ABC7384B8C075AE820090E113DBB4EABB22C>  
(letzter Zugriff: 22.01.2022)

*Bio*: <https://www.umweltbundesamt.de/umwelttipps-fuer-den-alltag/siegelkunde/bio-siegel-eu-deutschland>  
(letzter Zugriff: 21.01.2022)

*Fair Trade*: <https://www.fairtrade-deutschland.de/was-ist-fairtrade/fairtrade-siegel>  
(letzter Zugriff: 21.01.2022)





*Umweltplaketten und Umweltzone*: <https://www.umweltbundesamt.de/themen/luft/luftschadstoffe/feinstaub/umweltzonen-in-deutschland#5-wie-werden-umweltzonen-eingeteilt>  
(letzter Zugriff: 21.01.2022)

## DER BEGRIFF «NACHHALTIGER TOURISMUS» IM SOZIALEN, ÖKONOMISCHEN, ÖKOLOGISCHEN UND LINGUISTISCHEN BEREICH

*Isabel Gallego Gallardo, Department of French and English Philology, Cádiz University. Research Group HUM-485, Foreign language teaching: materials for a new curriculum design*

This article introduces the issue of sustainable tourism, highlighting the importance of its incorporation in all the institutions involved in the tourism sector, being an issue that still raises obstacles and questions on how to put it into practice.

First, the article defines the two terms “tourism” and “sustainable tourism”; followed by a differentiation of these two terms. Furthermore, a look at the evolution of these terms and the expansion throughout history is given. Moreover, the positive and negative impacts of tourism are being addressed, as well as the social and economic effects. The article concludes by exposing the lack of deference to the relationship regarding tourism and nature and its protection.

Analysing the tourism sector, one realizes the difficulties regarding the implementation of means and practices respecting the environment, to be able to integrate sustainability as a tourism project and tourism culture.

The presentation of the subject of ecolinguistics and the formation of German words and the development of ecological symbol has various objectives. Firstly, to emphasize the need to make it a priority topic in discussion forums, as this will enable proposals to be made regarding appropriate management and incorporation into the practices of the tourism sector. Secondly, to contribute through language to raising awareness of all those who habitually practice tourism without even asking themselves the question whether one day the world will cease to offer us what it has offered us up to now.

• *ecolinguistics* • *tourism* • *sustainability* • *sustainable tourism* • *ecology* • *environment*.

# Обучение немецкому языку как иностранному в тандеме с помощью приложения Hololingo!: анализ коммуникации в тандеме в игровой социальной виртуальной реальности

*Лазович Милица, доктор филологических наук, Университет Филиппа в Марбурге, Германия, milica.lazovic@uni-marburg.de*

*Алерс Тимо, доктор филологических наук, Потсдамский университет, Германия, timo.ahlers@uni-potsdam.de*

Изучение языка в тандеме в виртуальной реальности добавляет много инновационных аспектов в практику овладения иностранным языком. В статье представлен общий обзор практик социальной виртуальной реальности и их возможностей для обучения, а также описание приложения Hololingo!, разработанное нашей исследовательской группой, которое дает представление о тандемных взаимодействиях в контексте немецкого языка как второго. С точки зрения коммуникации мы сосредоточимся на некоторых аспектах действий в виртуальной среде, связанных, с одной стороны, с практическими действиями, а с другой стороны, речевым рефлексивным эпизодом игры, в котором действуют тандемные партнеры, для совместной интерпретации надписи в соответствии со своим языком и культурой. Количественное сравнение с тандемами *Hangoutt* (в *AltspaceVR*), а также оценка обратной связи с пользователем позволяют выявлять дополнительные потенциальные возможности игровых VR-приложений для будущей практики изучения языка в тандеме с использованием игр.

• немецкий язык как иностранный • изучение языка в тандеме • приложение Hololingo! • виртуальная реальность.

## 1. TANDEMLERNEN IN SOCIAL VIRTUAL REALITY

Die *Social-Virtual-Reality* (SVR) eröffnet Handlungsmöglichkeiten, die in einer neuen Form das Lernen in E-Tandems beleben. Fremdsprachliche Lernprozesse werden dabei aufgrund körperlicher Immersion, explorativen, aktionalen und kollaborativen Charakters und vielfältiger, uneingeschränkter Handlungsmöglichkeiten um neue Dimensionen erweitert. Ermöglicht werden multimodale Lernerfahrungen mit einem praktischen, lebensweltlichen Bezug. Im vorliegenden Beitrag untersuchen wir, inwiefern diese neue Praxis für den Erwerb diskursiver mündlicher Kompetenzen genutzt werden kann. Wie häufig empirisch gezeigt, wird der Erwerb mündlicher

Sprachkompetenzen durch bisherige Lernangebote oft als unzureichend wahrgenommen (vgl. Ahlers et al. 2020). Sprachtandems gelten allerdings als eine besonders beliebte und effektive Lernform (ebd.), denn sie ermöglichen den Lernenden im „direkten Kontakt mit einem Muttersprachler [oder fortgeschrittenem Sprecher], ihre Fremdsprachenkenntnisse wechselseitig anzuwenden, zu korrigieren und zu festigen“ (Bechtel 2010: 286). Für DaF-Lernende stellt aber das Bilden lokaler Tandems aufgrund von Verfügbarkeit, Sprachpassung, Sympathie, divergierender Interessen, Organisationsaufwand, fehlender Lernautonomie und Alltagsintegration oft eine Herausforderung dar. Daher werden zunehmend digitale, barrierearme E-Tandems genutzt (Funk et al. 2017). Allerdings kippen solche Audio- und Videotandems (z. B. *Hellotalk*, *Tandem*) wegen mangelnder gemeinsamer Aktivitäten schnell ins langweilige, vermittlungstypische Reden-um-zu-Reden (Ahlers et al 2020: 238). VR-Anwendungen hingegen ermöglichen durch eine spezielle stereoskopische ‚Brille‘ (*Head-Mounted Display*) sowie über Motion- und Handcontrollertracking das immersive, dreidimensionale Eintauchen in handlungsorientierte interaktive Welten, z. B. auch zum Zweck berufsbezogener Bildung (Zender et al 2019), welche den Handlungskontext dynamischer, erlebnisreicher, intensiver und authentischer gestalten. Während Einzelspieleranwendungen auf semiauthentische, geskriptete Kommunikationsszenarien mit Chatbots (*Mondly VR*: Getränk bestellen, im Hotel einchecken) oder auf immersives Vokabellernen beschränkt sind (*Word Saber*, vgl. Hartfill et al. 2020), stellen Mehrspieleranwendungen wie *AltSpaceVR*, *VRChat* oder *Rec Room* öffentliche und private Chaträume mit z. T. kurzweiligen Aktivitäten für freie, handlungsbasierte Zusammenkünfte bereit (Bowling, Tanzen, Schneeballschlacht). Wenige technisch affine Sprachlehrende bieten SVR-Sprachunterricht für Lerngruppen im Klassenzimmer oder als Exkursion über etablierte Social-VR-Plattformen (z. B. *AltSpaceVR*) an. Solche SVR-Umgebungen können auch zum autonomen Lernen genutzt werden.

Bereits 2D-Bildschirmmanwendungen (*SecondLife*, *TeCoLa*<sup>1</sup>) können authentische, barrierearme Redeanlässe in einer interaktiven, mehrsprachigen Sprachpraxis schaffen, interkulturelles Lernen begünstigen und Lernende zum Durchbrechen von Lernroutinen, zu kultur- und autoreflexiven Konstruktionsprozessen (vgl. Biebighäuser 2014: 369) anregen oder durch Erkundung unterschiedlicher Rollen in wechselnden sozialen Konstellationen durch Avatar-Nutzung fördern. Basiert auf dem Konzept des *Task-based-learning*s ermöglicht dies zielgerichtete, selbstbestimmte Partizipation in Kooperation mit einem lebensweltlichen Bezug. Auch in VR können solche Lernumgebungen etwa als konkrete (z. B. berufssprachliche) Zielszenarien realisiert werden (z. B. Lackierwerkstatt, Zender et al. 2019), die im Vergleich zu analogen Szenarien durch zusätzliche Funktionen wie digitale Hilfsmittel (Wörterbuch, Übersetzungstool, Notizen, Vokabel- und Grammatikinformationen) augmentiert werden können. VR könnte so empraktisches Problemlösungshandeln befördern und zu einer wichtigen

<sup>1</sup> Project: Pedagogical Differentiation through Telecollaboration and Gaming for Intercultural and Content Integrated Language Teaching (TeCoLa), <https://sites.google.com/site/tecolaproject/> [02.07.2021].

kognitiven Flexibilisierung führen sowie die soziale Kompetenz und Adaptivität durch emotionale Regulierung und Empathie fördern (vgl. *empathie machine* in Bertrand et al. 2018). Lernende werden in der VR ermutigt, ihre Handlungsfreiheit zu erkunden, sich in gewählten Rollen selbstregulativ zu entwickeln und die Wirksamkeit eigenen sprachlichen Handelns (vor allem Argumentation, Explikation, Elaboration) zu erleben (vgl. Mystakidis 2019).

Der mediale Mehrwert von VR liegt in der ausgeprägten Immersion. Die sensomotorische Kopplung des eigenen physischen Körpers an den virtuellen Avatar erfolgt durch: i) das Einnehmen der visuellen Egoperspektive, ii) die Steuerung der virtuellen Avatar-Gliedmaßen mittels korrespondierender physischer Kopf-Handgesten und Lokomotion sowie iii) sensorisches Feedback (haptisch erfahrbare Controller-Vibration, visuelle und auditive Eindrücke der Dreidimensionalität<sup>2</sup>) (vgl. Ahlers et al. 2021: 39). Wie beim analogen Tandem können auch im tele-ko-präsenten Distanzlernen mit VR (Zhao 2003) multimodale Ursache-Wirkungs-Zusammenhänge und Konsequenzen eigener empraktischer (Sprach-)Handlungen unmittelbar körperlich erfahren werden. Durch eine hohe Synchronizität physischer und avatarlicher Körperlichkeit, die virtuelle Erfahrungen in physischen, real-körperlichen Zuständen erfahrbar macht (*Cognitive Grounding*, Barsalou 2008), können kommunikative Praktiken und emotionale Prozesse real erfahren werden und tragen potenziell zu größeren Lerneffekten bei, etwa wenn zu einer Vokabel das ko-aktive motorische Konzept assoziiert wird (*Hebbian Learning*). Einige Studien zum Mikroerwerb im SVR-Kontext (vgl. Ahlers et al. 2020) zeigen wichtige Lernpotenziale bei der koordinativen Steuerung im virtuellen Handlungskontext für den Erwerb von Anaphern, Präpositionen und Diskursreferenzialität bzw. in der diskursiv-deiktischen Synchronisierung und im kommunikativen *Grounding* (Vandergrift 2016).

## 2. SVR-TANDEMABENTEUER: LERNEN MIT DER *HOLOLINGO!*-APP

Anders als smartphone-basierte Sprachlern-Apps, die meist geschlossene Aufgaben (Multiple Choice, Matching, Listen & Click) und diese oft als Pattern-Drill anbieten, können offene kollaborative, problemlösungsbasierte Aufgaben, Simulationen oder *Serious Games* komplexe diskursive Kompetenzen ansteuern (vgl. Blume/Schmidt/Schmidt 2017: 222, 227). *Digital-Game-Based-Language-Learning* (DGBLL) liefert durch die Einbindung solcher Aufgaben in narrative Abenteuer bereits in vielen Bereichen bessere Resultate als herkömmlicher Sprachunterricht im Klassenzimmer, etwa bzgl. Motivation, Lerndauer, Selbstwirksamkeit, Willens, in der L2 zu kommunizieren, bzgl. sozio-kommunikativer Kompetenzen, Engagements in begleitender schriftlicher Kommunikation (vgl. Hung et al. 2018). DGBLL-Tandems werden – anders als *Hangout*-Tandems – vom Auswählen, Ausdenken und sprachförderlichen Gestalten von Aktivitäten entlastet und durch einen narrativen Pfad zu kollaborativen, sprachdidaktisch förderlichen Aufgaben gelenkt, die sie frei bearbeiten können. Die Untersuchung der Nutzungspotenziale der VR-Chat-Plattform *AltspaceVR* für 1:1-Tandemlernen in Ahlers et al. (2020) zeigt (Abb. 1), dass die virtuellen Aktivitätsangebote in Tandemkonstellationen lernrelevante mündlich-diskursive Phänomene frequent evozieren können (etwa dynamisches *Turn-Taking*, korrekatives Feedback und viele andere syntaktisch und lexikalisch förderliche Lernkontexte).

<sup>2</sup> Künftige Hardwareprodukte versprechen umfassendes Bewegungstracking und haptisches Feedback (Teslasuit), Synthese von Geruch, Wärme, Wind (Feelreal Mask) und Mimiktracking (Mask).



Abb. 1: Eindrücke freizeitbasierter VR-Tandems: Zeigegestik (Kopf/Körper-Positur, Hand-Pointer) begleitet vom Deiktikon (*da*); multimodales Lernen durch Ko-Aktivierung von Motor- und Sprachkonzepten (z. B. das Verb *werfen* mit der motorischen Wurfhandlung)

Allerdings werden diese kaum strukturiert, als in ihrer Lernrelevanz bewusst erkannt, funktional eingeordnet und vertieft. Lernende werden infolge fehlender lernregulativer Kompetenz größtenteils überfordert, sind sich ihrer Funktion in den jeweiligen Tandemrollen wenig bewusst (vgl. ebd) und wünschen sich eine stärkere adaptive Begleitung, mehr Feedback in diesen Prozessen<sup>3</sup> sowie Anschlüsse zu anderen Lernkontexten. Nutzung des virtuellen Handlungs- als Lernkontextes sowie Transfer der geförderten Kompetenzen auf neue Handlungskontexte erweisen sich für Lernende als herausfordernd. Sie verlangen eine begleitende vor- und nachbereitende Sprachlernberatung und Hilfe bei der adaptiven Synchronisierung der unterschiedlichen Lernprozesse. Bis diese maschinell begleitet und im Gaming-Flow anhand von *Learning Analytics* angepasst werden können, bieten sich neben Sprachlernberatungen Game-integrierte Module zur Unterstützung in der Selbstregulation, im Transfer und in der Anpassung der Sprach-/Lern-prozesse an.

Mit der Game-Engine *Unity* haben wir daher die *Hololingo!*-App entwickelt (Ahlers et al. 2021), in der Tandempartner in einem Exit-Game-Szenario durch steuernde Aufgaben und spielerische, narrative Elemente zu mündlich-diskursiver Sprachpraxis angeregt werden (*nudging*). Die *Hololingo!*-App versucht dabei den Herausforderungen der Selbstregulierung durch eine optimale Koordinierung der authentischen situativen Lernoffenheit und der zielgerichteten, strukturierten Begleitung in sprachdidaktisch gestalteten, virtuellen Lernumgebungen zu begegnen. Dies geschieht u.a. durch die integrierte Förderung von Lernautonomie und Sprachreflexion: Die virtuelle Lernumgebung der *Hololingo!*-App enthält eine einführende Sequenz, mit dem Ziel der Bewusstmachung von Tandemrollen, -lernfunktionen und -abläufen sowie eine integrierte und nachbereitende Sequenz in der sprach- und metareflexive Prozesse angeregt werden. Ein weiterer Aspekt ist die Einbettung in ein narratives Abenteuer, das durch einen verbindenden Bogen die Tandems durch die Sequenzen zu einem Abschluss führt und motiviert. In den sprachdidaktisch aufbereiteten Lernsequenzen werden über kommunikativ-kollaborative Aufgaben und Sprachspiele (im Sinne von *Task-based Learning*) holistische Kommunikationsanlässe geschaffen, die auf die Synchronisierung unterschiedlicher Kompetenzbereiche abzielen. In weiteren

<sup>3</sup> Auch für Smartphone-basierte Sprachlern-Apps wünschen sich Lernende oft mehr Feedback zu Lernfortschritten und mehr Kontrolle bei der Einstellung des Schwierigkeitsgrades (Wu/Lee/Schmidt 2016).

Entwicklungszyklen der App sollen weitere Sprachspiele implementiert werden, die gezielt sprachliche Phänomene ansteuern, um eine optimierte Lernprogression zu ermöglichen: z. B. Sprachspiele zu Farb-Adjektiven, Tempus, Aspekt (Steels 2012) oder zu Relativsätzen (Ahlers 2018). Weitere Entwicklungsdesiderata betreffen den Ausbau von Hilfsmitteln (Scaffolding durch Bereitstellung von Redemitteln, Vokabeln oder einer Lernwortkartei), welche selbstbestimmt genutzt werden können und somit die individuelle Entwicklungszone ansteuerbar machen sowie das Tracking des Lernfortschritts durch kurze Selbsttests mit *Gamification*-System. Um den ästhetischen und Abenteuerereffekt zu intensivieren sowie interkulturelle Lernprozesse anzuregen wurde für *Hololingo!* eine Figur als Script-Träger ausgewählt, nämlich der Huckup, bekannt als Hildesheimer Sagengestalt, die (Apfel-)Dieben als schlechtes Gewissen metaphorisch ins Genick springt. Mit dem Spiel „Das Geheimnis des Huckup“ wurde dazu nun ein erster Demonstrator erstellt (vgl. dazu Abb. 2 und 3).

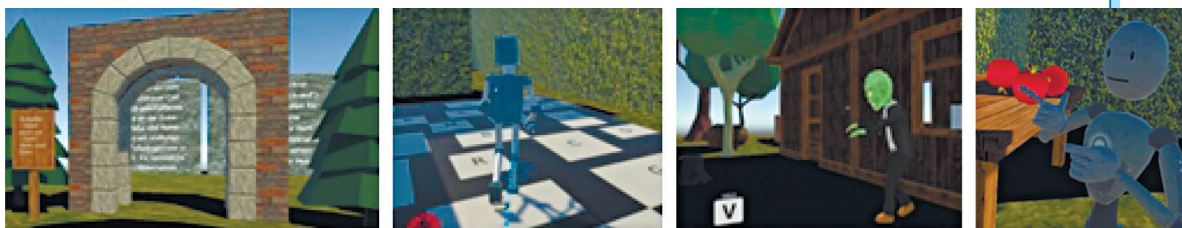
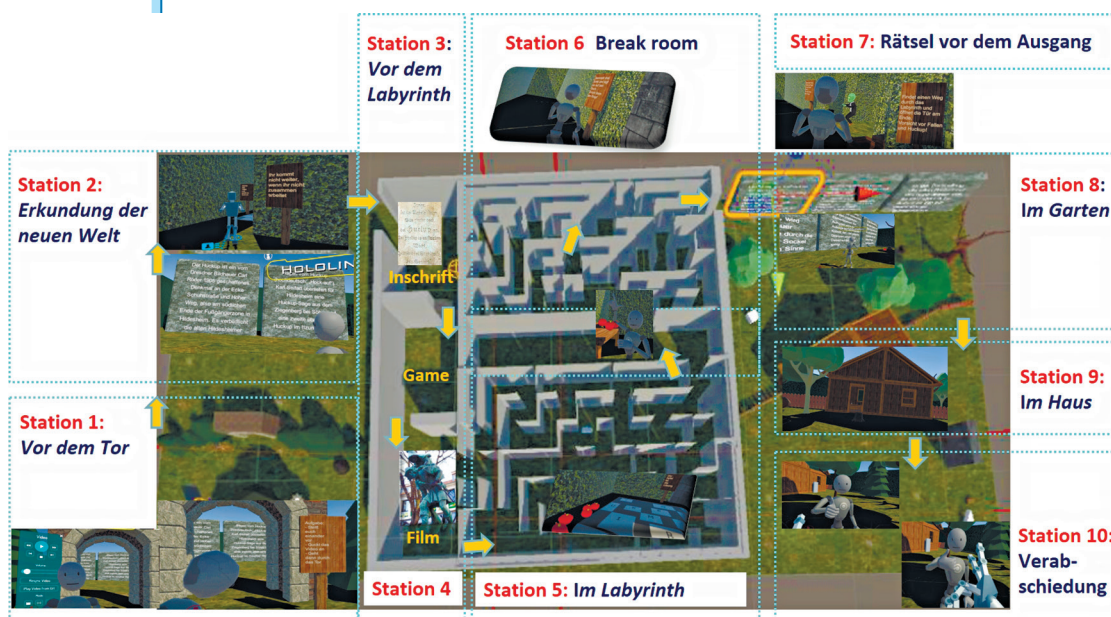


Abb. 2: Eindrücke aus der Egoperspektive: Eingang, Exit-Room-Sprachrätsel, Begegnung mit Huckup, Reflexionsphase

Im VR-Abenteuer *Hololingo!* geht es in verschiedenen Stationen darum, sich mit der Sage auseinanderzusetzen (etwa Übertragung einer Niederdeutschen Inschrift ins Standarddeutsche) und sie durch das Lösen von thematischen Exit-Room-Rätseln nachzuerleben. Die erste Station dient zum Kennenlernen der Tandempartner (TP); neben dem Smalltalk, der Gewöhnungs- und Synchronisierungsphase sehen sich die TP ein Erklärvideo zum Tandemlernen an, bestimmen ihre (Lern)Ziele und vereinbaren ihre Rollen und Aufgaben für die anstehende Tandemarbeit. Sie lernen anschließend die Aufgabenstellung kennen und erkunden das Setting (Station 2). Das Lernabenteuer beginnt explorativ und multimedial: Die TP sollen sich zuerst die Legende über einen Stummfilmausschnitt und die gemeinsame Übersetzung der niederdeutschen Inschrift der Huckup-Statue erschließen oder in einer sprach-basierten Gaming-Sequenz zur Öffnung einer Tür koordinieren (Station 3), um den Inhalt der Sage interpretativ einzuordnen und reflektierend zu diskutieren. Danach erfolgt eine dynamische Episode im Labyrinth, wo sie kollaborativ Exit-Room-Aufgaben lösen, wie etwa Buchstabenrätsel, bei denen das Lösungswort mit Äpfeln auf Steinplatten markiert oder anhand von im Labyrinth versteckten Hinweisen in ein Buchstabenschloss eingegeben werden muss. Dabei dürfen sich die TP nicht vom Huckup fangen lassen, müssen sich gegenseitig warnen und kollaborativ navigieren (Stationen 4 und 6). Zwischendurch und nach dem Labyrinth gibt es Gelegenheiten für Pausen, die dank eingebauter Gesprächsimpulse für Smalltalk, zur Reflexion, Verständnissicherung und Planung nächster Schritte genutzt werden (Stationen 5 und 7). Es folgen eine spontane Erzählung und Austausch über eine vergleichbare Sage aus dem Sprach-/Kulturraum der L2-Lernenden (Station 8) sowie sprachkontrastive Übungen (z.B. beim Vergleich von Sprichwörtern in ihren Ausgangssprachen). Abschließend finden ein Feedbackgespräch zur Lernprogression und zur Tandemarbeit sowie die Verabschiedung statt (Station 9) (für eine ausführliche Darstellung vgl. Ahlers et al. 2021).

Abb. 3: Schematischer Aufbau des *Hololingo!*-Abenteuers

Obwohl uns zur Weiterentwicklung der App im aktuellen Stadium primär qualitative Analysen interessierten (s. folgende Kap.), wollten wir in einem ersten Schritt wissen, ob sich das Design der App auch quantitativ auf die Tandem-Kommunikation auswirkt. Für eine erste Annäherung verglichen wir zehn *Hangout*-Tandem-Aufnahmen in der Social VR Anwendung *AltspaceVR* (Gesamtdauer 3 h 53 Min.) mit einer längeren zweiteiligen Tandem-Aufnahme in der tandemspezifischen App *Hololingo!* bezüglich der Gesprächsbeteiligung und Diskursanteile der Tandempartner. Als Grundlage zogen wir die Gesprächsdauer und Gesamtwortzahl (inkl. Diskurspartikel wie *äh, mh*) heran sowie die Anteile der Gesprächspartner daran (s. Tab. 1). Die Tandems bestanden jeweils aus L1-Experten und L2-Lernenden des Deutschen mit unterschiedlichen Erstsprachen, die als Erasmusstudierende ihr Auslandssemester in Deutschland verbrachten.

Tab. 1

#### Vergleich der Diskursaktivität von *Hangout*-Tandems in *AltspaceVR* mit *Hololingo!*<sup>4</sup>

Aufnahme	Tandem-paare	Dauer (Summe)	Wortzahl (gesamt)	Wortzahl/ Min.	Wortzahl (L1)	Wortzahl (L2)	Anteil Wortzahl (L2/gesamt)
<i>Altspace VR</i>	10	3:54:52	14009	59,6	9534	4475	0,32
<i>Hololingo!</i>	1	1:27:28	4072	46,6	2206	1866	0,46

Die Analyse der Diskursaktivität zeigt, dass der Gesprächsanteil der L2-Lernenden (mit 46%) in der *Hololingo!*-App tendenziell höher ist bzw. die Tandemarbeit symmetrischer und kommunikativ ausgeglichener verläuft als bei *Hangout*-Tandems in *AltspaceVR* (mit 32%), die damit asymmetrischer ausfallen. Offene Spielwelten ohne sprachdidaktische angepasste Angebote scheinen

4 Der Zusammenhang zwischen VR-Umgebung (*AltspaceVR/Hololingo!*) und Tandemrolle (L1/L2) bzgl. der Wortzahl ist signifikant:  $\chi^2(1, N=18081)=266,98, p=,000, \phi=0,12$ .

die Tandempartner also nicht so gut dabei zu unterstützen, an der Kommunikation teilzunehmen. Die L1-Experten ergreifen häufiger die Initiative, um den Gesprächsfluss aufrecht zu erhalten, was zur Intensivierung der kommunikativen Asymmetrie führt, aber auch zu interaktiven Problemen in der Bestimmung gemeinsamer Ziele und bei der Eröffnung neuer Handlungskontexte. In der *Hololingo!*-Lernwelt konnte dagegen die L2-Lernende durch Angebote der Lernumgebung die Selbststeuerung besser übernehmen. Der Vergleich der interaktiven Beteiligung der L1- und L2-Sprecherinnen in unterschiedlichen Stationen der *Hololingo!*-App bestätigen zudem interaktive Dynamiken in unterschiedlichen Handlungskontexten: So ist die L1-Expertin bei der eher unterrichtsnahen Übersetzungsaufgabe mit 67% Gesprächsanteil zwar dominanter, aber meist in unterstützender Funktion durch Hinweise und Initiierung inferenzieller Stränge für die L2-Tandempartnerin. In kommunikativen Problemlösungsaufgaben, etwa beim Lösen eines kooperativen Exit-Room-Rätsels (Station 4) konnte die L2-Lernerin mit 46% Gesprächsanteil ausgeglichen agieren. Bei der reflexiven Station verzeichnete L2 beim gegenseitigen Austausch über den Lernprozess mit 54% ihren höchsten Gesprächsanteil. Die lernförderlichen Dynamiken in der Gesprächsbeteiligung und die Vielfalt unterschiedlicher Handlungskontexte, in denen Lernende sukzessive in der Selbständigkeit wachsen, deuten einen weiteren Vorteil dieser Lernumgebung im Vergleich zu *Hangout*-Tandems in SVR an. Eine umfassendere Folgestudie mit weiteren Aufnahmen müsste die Tendenz der vorliegenden Ergebnisse prüfen.

### 3. MEINUNG DER LERNENDEN ZUM HANDLUNGSRAUM UND LERNPROZESS IN DER HOLOLINGO!-APPS — „ICH HATTE DAS GEFÜHL, NEUE WÖRTER MENTAL TIEFER ZU VERANKERN, DENN MAN SPRICHT UND HANDELT GLEICHZEITIG“

Nach der Erprobung des Lernszenarios der *Hololingo!*-App wurde im Nutzer-Feedback eine ausgeprägte Zufriedenheit sowohl mit dem Prozess als auch dem Lernergebnis festgestellt. Das explorative, erlebnisreiche, neue Setting wird bereits beim aktuellen vorläufigen Entwicklungsstand besonders positiv hervorgehoben. Gleiches gilt für die kognitive Prägnanz der neuen Lernerfahrung, die Möglichkeit der kollaborativen Bestimmung der Ziele und genauso die Koordinierung im gemeinsamen Handeln und Wahrnehmen sowie in der geteilten Expressivität, die als besonders wichtiger interaktiver und *Ingrouping*-Moment erlebt wird. Es bestätigt sich, dass der Kontext positive Auswirkungen auf die Motivationssteigerung, die Handlungsorientierung, die eingeschätzte Selbstwirksamkeit und das Gefühl der sozialen Nähe hat. Die Probanden heben die Möglichkeit hervor, als kulturelle Vermittler zu agieren und besonders intensive interkulturelle Erfahrungen zu machen.

„Man hat das Gefühl etwas Außergewöhnliches erlebt und bewältigt zu haben. Das gemeinsame Arbeiten und Lösen von Aufgaben hilft beim Kennenlernen. Diese Art des Kennenlernens und interkulturellen Kommunizierens ist sehr angenehm und abwechslungsreich.“ (Auszug aus einem anschließenden Interview mit L2-Lernenden)

Während die Lernende mit Deutsch als Fremdsprache das erlebnisbasierte Handeln und die geförderte Expressivität sowie das emotionale Handeln verstärkt im Lernprozess spürt und das Gefühl hat, mehr mitgenommen, verstanden und integriert zu werden bzw. durch perzeptive, emotionale und aktionale Synchronisierungsaktivitäten das Soziale näher und intensiver zu erleben, hebt die Probandin mit Deutsch als L1 die Gegenseitigkeit im interkulturellen Austausch und die gemeinsame Reflexion zentral hervor. Sie wünscht sich sprachlich noch mehr auf Interaktivität angelegte Aufgabenstellungen, welche die interaktive Harmonie befördern und den gegenseitigen, interkulturellen Austausch





auf der gleichen Augenhöhe ermöglichen (z.B. durch Sprachwechsel und sprach- bzw. kulturkontrastierende Impulse).

Beide DaF-Lernenden loben die authentischen, dynamischen Handlungsmöglichkeiten zum Ausbrechen aus dem musterbasierten Handeln und Anwenden von auswendig gelernten oder gedrillten Formen. Lernende betonen die stark wahrgenommene Selbstwirksamkeit sowie die Möglichkeit, unbekannte Wörter aus dem authentischen kommunikativen Bedürfnis bei der Aufgabenrealisierung integrativ zu lernen, interaktiv auszuhandeln und in der Anwendung direkt zu erproben. Lernende kommentieren: „*Man lernt direkter die andere Denkweise kennen und anzueignen*“ und „*Man bekommt die Möglichkeit, neue Wörter mental tiefer zu verankern, denn man spricht die Wörter selbst und handelt gleichzeitig*“. Die neue körperliche Erfahrung und Selbstwahrnehmung sowie das Lernen mit hoher emotionaler, affektiver Involvierung sorgen dafür, dass die Lernenden Wortschatz funktional und kognitiv integrierter erfahren und somit fester verankern, weswegen dieser auch als erinnerungsstabiler wahrgenommen wird. Diese positive Selbstwahrnehmung der „tieferen“ Lernqualität sollte aber durch objektivere Verfahren noch geprüft werden.

Trotz vieler positiver Aspekte schlagen die Tandempartner einige Optimierungen im Setting vor, bedingt unter anderem durch Erwartungen an Lernroutinen aus traditionellen Unterrichtskontexten: So wünscht sich z.B. die DaF-Lernende mehr integrierte Aufgaben im „klassischen Sinne“ (des *Focus-On-Form*), mehr Komponentenübungen sowie zur bewussteren Einübung neuer Chunks, Lückentextaufgaben und Wortschatzübungen sowie ergänzende „*take-home-Materialien*“, in denen die Wörter, Strukturen oder Texte außerhalb des virtuellen Settings bewusster und vertiefter bearbeitet werden könnten. Sie wünschen sich auch Unterstützung in der Bestimmung der Lernziele und im *Designing* der Lernprozesse sowie korrekatives Feedback und die funktionale Möglichkeit, leichter zu anderen Episoden zu wechseln, also abwechslungsreichere Navigations- und Exit-Strategien. Dieser Übergang zu einer neuen Lernqualität mit sukzessive zu entwickelnder Selbstregulation und ergänzenden Nutzungsformen wäre bei der Konzeption und Etablierung neuer Lernsettings zu beachten. Diese könnten z. B. in Form von vor- und nachbereitender Aktivitäten, als Lernbegleitung oder digitales Sprachlerncoaching zum Einsatz kommen. Auch wären eingebaute Selbsthilfe-Buttons denkbar, über die Lernende zusätzliche Materialien oder Multimedia integrieren können. Über den Einbau von Tests könnten Übergänge und Progressionsrouten individuell angepasst werden, um selbstgesteuerter und -bewusster die eigenen Lernprozesse mit höherer wahrgenommener Handlungsmacht konstruieren zu können.

#### **4. ANALYSE DER KOMMUNIKATION IM HOLOLINGO!-TANDEM**

Die ersten quantitativen Analysen der Diskursanteile (vgl. Kapitel 2) haben gezeigt, dass die *Hololingo!*-App eine intensivere, ausgewogenere und partizipativere Interaktion im Vergleich zu *Hangout*-Tandems ermöglicht. Genauso haben die Tandempartner (vgl. Kapitel 3) viele subjektiv

wahrnehmbaren Vorteile für die Lernprozesse genannt. Im vorliegenden Kapitel möchten wir dies interaktionsanalytisch vertiefen und einen mikroskopischen Einblick in die erwerbsförderlichen Prozesse dieses interaktiven Settings geben. Dabei fokussieren wir einerseits die sprachreflexive Episode mit der kollaborativen Bearbeitung der Sprachaufgabe zur Entschlüsselung einer niederdeutschen Inschrift und schauen auf das sprachliche Unterstützungshandeln und seine lernförderlichen Potenziale (vgl. Kapitel 4.2.). Andererseits wird die Station im Breakout-Room (vgl. Kapitel 4.3.) näher betrachtet, wo sich das kultur- aus dem sprachreflexiven Handeln ergibt. Da das empraktische, emotionale und expressive Handeln in der virtuellen Umgebung besondere Relevanz hat, nehmen wir diese als Ausgangspunkt unserer Analyse (Kapitel 4.1.). Als Datengrundlage dient hier exemplarisch die chronologisch jüngste Tandeminteraktion, die im Rahmen der iterativen Entwicklung der *Hololingo!*-App aufgenommen wurde und aktuell den letzten Entwicklungsstand der App repräsentiert. Die Aufnahme umfasst insgesamt 87 Minuten einer authentischen Tandeminteraktion mit einer deutschen Studentin und einer (ihr unbekannt) DaF-Lernerin mit Chinesisch als L1, die sich gerade auf dem B2-Lernniveau (nach GER) befindet. Relevante Interaktionsausschnitte wurden in HIAT transkribiert und mittels multimodaler Kommunikationsanalyse ausgewertet (Ehlich 2007; Redder/ Rehbein 1999).

#### 4.1 RESSOURCEN DES EMPRAKTISCHEN UND EXPRESSIVEN HANDELNS IM TANDEM

Die interaktiven Potenziale des Settings liegen nicht nur in der geteilten Perspektivierung, der Handlungskoordination, Synchronisierungsaktivitäten und kooperativen Lösungsfindung, sondern in der spielerischen Erprobung der neuen körperlichen Realität mit neuen interaktiven Möglichkeiten. So spielen die Tandempartner (TP) mit dem Körper des Avatars, mit der Selbst- und Fremdwahrnehmung sowie mit den neuen Handlungsmöglichkeiten des virtuellen Raums, springen und fliegen. Sie spielen mit Schatten und kommentieren die virtuellen „Naturphänomene“ (s. Abb. 4). Sie teilen auch virtuelle Herzchen oder Smileys aus oder kommentieren ihre körperliche Wahrnehmung besonders expressiv, die dann mit viel Expressivität und Emotionalität geladen für interaktive Höhepunkte sorgen: z.B. *Oh, was passiert? Ich bin ganz (•••) / Ich bin ganz in der [sic!] Himmel! (Beide lachen).*

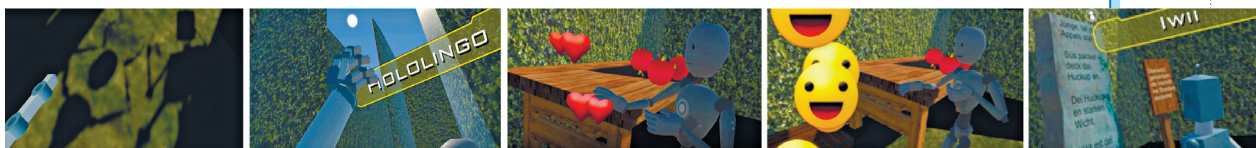


Abb. 4: Schattenspiel, Zeigegeste, Emoji-Einsatz, gemeinsame Blickausrichtung

Da der virtuelle Raum viele Freiheiten sowie Unsicherheiten bzgl. Positionierung, Bewegungsmöglichkeiten und Orientierung bringt, sind die Tandempartner nicht nur kognitiv und aktional aufeinander angewiesen und um möglichst maximale Synchronisierung in der Wahrnehmung, Expression und Aktion bemüht, sondern sie scheinen auch ein stärkeres Bedürfnis nach der „körperlichen“ Nähe, affektiven und interaktiven Sicherheit zu verspüren. Hierfür stehen häufige Äußerungen wie „*ich komme dir ein bisschen näher*“ und solche zur Sicherung der interaktiven Verbundenheit in der Wahrnehmung und im Handeln. Die geteilte körperliche Erfahrung, das expressive und emotionale Kommentieren führen zur Verschmelzung der emotionalen

Horizonte und für DaF-Lernende zum intensivieren emotionalen Handelns in der Fremdsprache. Zumal einige nonverbale Elemente fehlen (etwa fehlende Gesichtsexpressivität der Avatare), wird eine Fokussierung auf die prosodischen Cues ermöglicht bzw. der Erwerb der prosodischen Pragmasemantik begünstigt. Des Öfteren werden hierzu Aussprache-Spielereien beobachtet, in denen sich die TP gegenseitig nachahmen, gewisse prosodische Elemente verzerrt oder unterhaltend übertrieben darstellen, aber auch Akkommodationsprozesse (vgl. Giles 2016) als Quelle zur Navigation im virtuellen Raum.

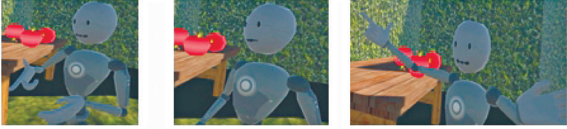
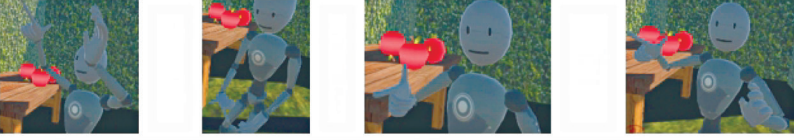
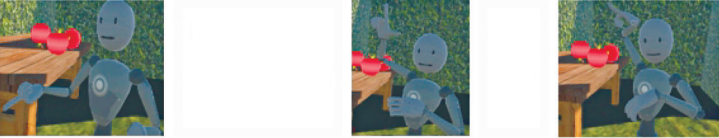
Auch wenn die Interaktion über Avatare erfolgt, was die Möglichkeiten der Nutzung von Mimik reduziert, wird die Gestik verstärkt genutzt. Die gestischen Ressourcen führen i) zu einer anderen Art kognitiver Aktivierung und Dynamisierung bzw. zu einer intensiveren In-Bezug-Setzung des Körperlichen mit dem Aktional-Interaktionalen, ii) zu einer multiplen Vernetzung der kognitiven Infrastrukturen (intern oder mit neuen Elementen) sowie iii) zu einem Neu-Erleben und empraktischen Restrukturieren bekannter Strukturen. Eine Intensivierung der Gestik bei L2-Lernenden wird besonders bei bekannten Wörtern mit einer spatialen Komponente (Präpositionen, Beschreibung der Form/Größe, Verben mit der Semantik der Lozierung/Bewegung, Deiktika) beobachtet. Die gesteigerte Gestik erfolgt als interaktive Rekonstruktion der Räumlichkeit in der VR bei der Synchronisierung der körperlichen Wahrnehmung, aber genauso bei der Verdeutlichung von syntaktischen Relationen sowie bei der Unterstützung in der Kohärenzherstellung (z.B. bei Bedingungssätzen, bei adversativen Konjunktionen oder bei Anaphern). Auch bei epistemisch relativierenden, Vagheit indizierenden Sprechhandlungen und in allen wichtigen Bereichen für die Förderung der Hörerorientierung, tritt ebenfalls intensivierte Gestik auf. Im folgenden Beispiel kommt dies sehr deutlich zum Ausdruck.

### BEISPIEL 1<sup>5</sup>:

So wie das Beispiel zeigt, wird in der Formulierung der lozierenden und explizierenden Elemente (*ein Ring / ganz groß / auf dem / auf dem Kirche / stehen – stellen*) eine besonders intensive, veranschaulichende Gestik verwendet. Dies kann auch in Fläche 2 bei den interaktiv-diskursiv strukturierenden Elementen ja, also ja und in Fläche 3 bei epistemischen Elementen *Ja, vielleicht* festgestellt werden.

Eine beobachtete Kuriosität zum *Embodiment*, die durch weitere Studiensicherlich zu überprüfen wäre, bezieht sich auf die Bewegungen der Selbstregulation (*moves of self-regulation*). Der Körper des Avatars ermöglicht eine neue Art der Selbstwahrnehmung durch Selbst-Distanzierung, die trotz aller kognitiver Herausforderungen für das Selbstmonitoring im Handeln und im Erwerbsprozess sehr förderlich zu sein scheint. Die intensivierte

5 Die Partitur-Transkripte enthalten neben den Äußerungen der Tandempartnerin (TP) mit Deutsch als Muttersprache (L1) und der TP mit Deutsch als Fremdsprache (L2) auch bildliche Darstellungen der Bewegung des sprechenden Avatars, aus der Perspektive der anderen TP. Dabei wird in der Beschreibung auf die Bewegungen der Hände, der Arme und des Kopfes eingegangen.

<p>1</p>	<p>L2 [v] es war ein Dieb und es / er möchte ein ähe / wie heißt das? ein ein Ring oder soein</p> 
<p>2</p>	<p>L1 [v] Ein Ring? Eine Glocke? Eine Glocke? L2 [v] ja, also ja ähe so wie das und äh es kling es klingelt. Ganz groß. Manchmal</p> 
<p>3</p>	<p>L2 [v] kann man /Ja, vielleicht. Also manchmal auf dem kann man auf dem Kirche sehen.</p> 

Gestik kommt hier bei der L2-Lernerin in einer spezifischen Funktion, nämlich bei der reflektierten Anwendung des regelbasierten Wissens und in der nonverbalen Selbstregulation vor. So nutzt die L2-Lernerin im folgenden Beispiel die Gestik des Avatars einerseits als *mental display* zur Synchronisierung mit der Tandempartnerin und als Exothese zu inneren mentalen Prozessen, also auch koordinativ und diskursiv-regulativ. Andererseits ermöglicht dies eine erwerbsförderliche Neustrukturierung bei der selbstreflektierenden, bewussten Anwendung gewisser bereits erworbenen Regeln. Hier zeigt sich eine interaktiv bedingte Dynamik der Selbstregulation, die dazu durch die neue körperliche Erfahrung im SVR eine zusätzliche kognitive und regulative Qualität bei der Verknüpfung einer erworbenen Regel mit Anwendungskontexten aufweist. So unternimmt die L2-Lernerin im folgenden Beispiel eine Abwägung potenziell passender Formen für das Verb *stehlen*. Sie verwendet bei jedem Token (vgl. Transkript: *stiehlt – stahl – stehlt – stahl*) eine andere Geste, die eine eigene konzeptionelle Qualität verkörpert bzw. eine empraktisch begleitete mentale Restrukturierung andeutet und interaktiv auf Suchprozesse und Bestätigungsappelle hinweist.

## BEISPIEL 2:

Anschließend wird die Geste durch maximale Öffnung aufgelöst und mit der Kopfhebung nach oben kombiniert, mit der parallelen Benennung der übergeordneten grammatischen Kategorie (vgl. in der Fläche 1 *ja, Konjugation*), als Indikator der Aktivierung des regelbasierten Wissens und einer Analogie-Operation, die hier mit einer ikonischen Geste realisiert wird. Diese zeigt sich in der Fläche 2 (vgl. *wenn – dann so*): Im Bedingungssatz wird eine typische ikonische Bewegung (*hin/oben/Regel => her/unten/Beispiel*) verwendet und als Indikator einer Analogieoperation richtig formuliert (er stahl). Die Gestik wird als Display und zur interaktiven Synchronisierung genutzt, aber auch als selbstreflexive Operation, die für die interne, konzeptionelle, anwendungsorientierte Ausdifferenzierung förderlich ist. Nach dem falsch interpretierten Hörsignal (*Jaa?*) wird allerdings diese Lösung erneut interaktiv angepasst (*er stahlte*) und dabei eine falsche, hybride Struktur mit zwei kollidierenden Regeln generiert (*stahlte*). Die Gestik dient auch hier als Indikator

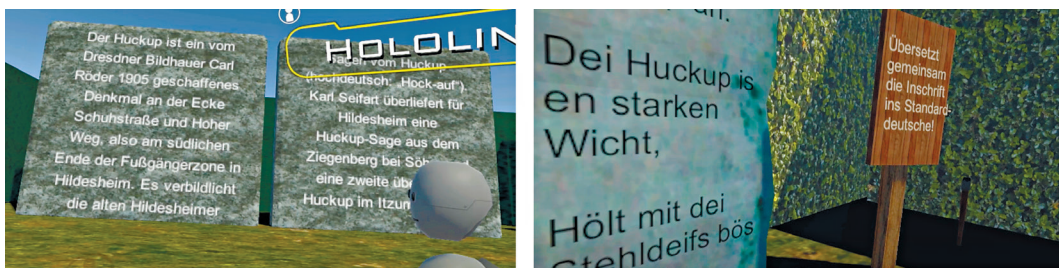
	L1 [v] nicht.		Mhm.
	L2 [v] Ja, aberer danach also trägt also stiehlt stahl		steht, ja, Konjungation.
1			
	L1 [v]	Jaa?	Ja.
	L2 [v] Wenn es in Vergangenheit so: Er stahl.		Er stahlte dann Glocken. Er hört nichts, also er
2			
	L2 [v] findet, wenn er nichts hört, dann all alle anderen auch nichts hören.		
3			

eines mentalen Prozesses, des hypothetischen Neuversuchs und markiert auch die Öffnung der interaktiven Aushandlungssequenz (s. die offene Geste mit beiden ausgestreckten, gedrehten Zeigefingern). Direkt im Anschluss (Fläche 3) wird mit der gleichen Verwendung des Bedingungssatzes (*wenn – dann*) eine ähnliche gestische Dynamik wie vorher beobachtet: Mit den ausgestreckten Zeigefingern und der Bewegung von oben nach unten wird intern ein Abgleich der Regel in der Anwendung gemacht und so eine selbstreflexive Handlung praktiziert, die gleichzeitig einen interaktiven Charakter der Synchronisierung hat. Dieses gestische Muster wird konstant mit der Bewegung des Kopfes begleitet. Die Gestik wird hier konsequent als Stütze in der konzeptionellen Ausdifferenzierung zwischen unterschiedlichen grammatischen Strukturen genutzt, in der Neurelationierung der Regel mit der Regelanwendung und zugleich als Display zur hörerorientierten Synchronisierung, zur Öffnung/ Schließung des interaktiven Raums und in der diskursiven Strukturierung. Dieses (durch die neue kognitive VR-Realität bedingte) intensiviere körperliche Handeln mit einer neuen interaktiven Erlebnisqualität und Selbstwahrnehmung beeinflusst die kognitive Dynamik bei der Regelaktivierung. Es regt zum Selbst-Monitoring an und ermöglicht eine neue, multimodal bereicherte Anwendung der aktivierten Regel. Die hier beobachteten *moves of selfregulation* sind nicht nur ein Beleg des medial geförderten Selbstmonitorings, sondern auch der intensivierten, interaktiv-kommunikativ ausgerichteten *Reflection-in-action* bzw. der interaktiven Regulation durch das Nonverbale, das sich als relevant für das Synchronisieren im mentalen, aktionalen und interaktionalen Bereich mit dem Gesprächspartner darstellt.

#### 4.2 Kollaborative Deutungsprozesse in der sprachreflexiven Aufgabe „Geheime Inschrift entschlüsseln“

Das Ziel dieser Gaming-Episode ist die kooperative Arbeit an der Entschlüsselung einer Inschrift, verfasst im Hildesheimer Platt, ins Standarddeutsche.

Dadurch soll die eingangs auf den Steinstelen eingeführte Sage, sprachlich und kulturell stärker reflektiert, kommentiert und als Gesprächsimpuls genutzt werden. Es handelt sich um keine einfache textbasierte Übersetzungsaufgabe, jedoch können die TP außerhalb des Textes explorativ handeln und den einleitend-erklärenden Text aus der vorherigen Szene sowie einen kurzen Film aus der nächsten Szene für die Erschließung nutzen, der die Sage des Huckups ästhetisch und narrativ darstellt. Auch lassen sich andere Gegenstände und *Gaming*-Szenen im späteren Verlauf erkunden. Das Setting ist mit vielen unterschiedlichen kontextuellen und multimodalen Cues bereichert, die zu kooperativer Exploration, Argumentation und *joint reasoning* einladen und somit den Charakter eines Lernabenteuers bei der sprachreflexiven Deutung der „geheimen Inschrift“ unterstreichen.



Inschrift (rechtes Bild): *Junge lat dei Appels stahn, Süs packet deck dei Huckup an, Dei Huckup is en starken Wicht, Hölt mit dei Stehldeifs bös Gericht.*

Abb. 5: Der Auszug aus dem Gaming-Kontext der Übersetzungsaufgabe

Die kooperative Arbeit an der verständnisorientierten Cues Entschlüsselung und anschließenden sprachlichen Formulierung und Anpassung fördert sprachreflexive und inferenzielle Kompetenzen. Dazu werden multilinguale, multimodale Ressourcen genutzt und es ist eine Synchronisierung unterschiedlicher inferenzieller Hinweise erkennbar. Der explorativ-spielerische Modus ermöglicht es, dass die beteiligte L2-Lernende kreativ, hypothesengenerierend, symmetrisch, ohne affektiven Druck und in einem gesichtswahrenden Rahmen agiert, dabei aber passende Unterstützung bekommt, die nicht reparativ, sondern als kooperative Wissensgenerierung im auf-/ausbauenden Modus erfolgt. Die Lernende gibt Impulse interaktiv symmetrisch, begründet diese, passt diese an und bemüht sich um kohärente Übergänge, nicht nur lokal, zwischen den Zeilen und interpretativen Blöcken, sondern auch diskursiv makroskopisch, in der Verknüpfung späterer Episoden im Spiel. Sie lernt, neue Elemente zu kontextualisieren und in der Fremdsprache logisch zu denken und zu argumentieren. L2-TP initiiert selbstregulativ kleinere erklärende Nebensequenzen, um Intersubjektivität und gegenseitiges Verständnis zu sichern und partizipiert auf eine neue Weise in der kooperativen, zielgerichteten Wissenskonstruktion, was auch Vandergrift (2016) in ihrer Studie als besonders wichtig hervorhebt: „*This process [...] helps the learner internalize knowledge [...] through attaining intersubjectivity, mutual understanding [...] and co-construction of knowledge*“ (ebd. 186). Die TP synchronisieren sich bei der Deutung und in der interpretativen Einordnung einzelner Elemente in der Nutzung lokaler, induktiver *bottom-up*- und kontextuell bedingter und *top-down*-Verarbeitungsprozesse. Sie unterstützen diese durch Verständnissicherungen, Begründungen und Kontextualisierungen und trainieren somit *Grounding* und *joint understanding* in einem interkulturellen Kontext. Unterschiedliche Lesarten in Verbindung mit der Arbeit an dissonanten Lösungsversuchen evokieren nicht nur



eine geteilte Verständigungsbasis, sondern sensibilisieren für die Prozesse mentaler und interaktiver Synchronisierung. Auch fördern sie Reflexion und Anpassungskompetenz bzgl. der oft latent differenten Inferenzquellen und Deutungsmuster. Um dies zu veranschaulichen, möchten wir das vorliegende Beispiel präsentieren, in dem sich das initiative, symmetrische, hypothesentestende Handeln der L2-Lernerin und das kooperativ-ausbauende und lernförderliche Fortführen der Tandempartnerin mit Deutsch als L1 zeigen: Beide bringen eigene Vorschläge stets als mögliche, kooperativ zu überprüfende Lösungen ein und agieren somit kooperativ-ausbauend und synchronisierend-aushandelnd.

### BEISPIEL 3:

Dabei zeigt sich, dass die Tandempartnerinnen unterschiedliche *Cues* in ihrer inferenziellen Aktivität nutzen: Während die L2-Interaktantin vor allem prosodische und semantisch-basierte Inferenzen nutzt, Elemente isoliert, diese fragmentiert betrachtet und Vorschläge eher weniger begründet, nutzt die Tandempartnerin mit Deutsch als L1 eher syntaktische *Cues*, begründet und kontextualisiert eigene Deutungsversuche und nutzt mehrsprachige *Cues* sowie jene aus dem Eingangstext und dem aktivierten Script. Somit unterstützt die L1-Sprecherin die L2-Sprecherin in der Orchestrierung unterschiedlicher inferenzieller Prozesse, wobei dies im Modus einer kooperativen, kontextualisierenden Fortführung erfolgt. Lösungsvorschläge der L1-TP werden stets mit Relativierungen, Vagheitsindikatoren, Approximativ-Indikatoren und einem Angebot zur kooperativen Validierung verbunden. Dies stärkt die Aktivierung und die Integration der L2-Sprecherin in der Phase der Lösungsfindung und somit ihre Selbstwirksamkeit, Handlungsorientierung und Integrität. Die Lösung erfolgt somit immer als *joint action*. Außerdem führt der interaktive Rahmen zu kurzfristigen Erwerbsfortschritten: Die Tandempartnerin mit Deutsch als L2 wird nicht nur aktiver, flüssiger und freier im Gespräch, sondern übernimmt (das Modell der L1-Sprecherin spiegelnd) einige interaktive Strategien (wie im Kontext der Begründung, Verständnissicherung und Selbstrelativierung oder bei der kooperativen Initiierung zur Überprüfung eines Vorschlags). Auch übernimmt sie die ko-konstruktiv entwickelten, inferenziellen Strategien (wie Kontextualisierungen und die Nutzung von syntaktischen *Cues*).

Die L2-Sprecherin handelt zudem expliziter und direkter bei Wortschatz-Problemen oder bei der Überprüfung eigener Hypothesen, durch direkte Nachfrage. Sie initiiert Erklär- und Nebensequenzen zur Klärung wichtiger Fragen. Die L2-Sprecherin agiert direkter und selbstinitiativer bei der Initiierung und Regulierung der Übergänge zwischen den Modi der Vermittlung (Erklärung oder reparative Nebensequenz) und der kooperativer Lösungsarbeit. Die L1-Sprecherin greift hingegen nur bei expliziten Aufforderungen explikativ-korrektiv ein. Sie handelt supportiv durch Strategien, wie symmetrische Fortführung, Einbringen weiterer Vorschläge oder Wiederholung mit einem *Cue*, um so die Lösung als gemeinsame Leistung vorzubereiten. L1 unterstützt auch

[5]	L2 [v]		Junge, lat/Dei steht für dein?
[6]	L1 [v]	Ähm, Würde ich auch sogn.	(• •) Appels (•) vielleicht (•) also
	L2 [v]	(•)	Ähm (•) Appels stahn.
[7]	L1 [v]	aus dem	aus dem Englischen apple, also. Wie jaa (lachen), vielleicht
	L2 [v]	mhm.	(•) Apple?
[8]	L1 [v]	apple?	Junge / und lat, was würdest du sagen lat?
	L2 [v]	Vielleicht.Aha.	(•••) Ähmmm,
[9]	L1 [v]		lat / Junge / aber, (2s) Ähm
	L2 [v]	Lat la wie einladen?Laden (•) ein?(•••) la (•••)	Ähm?
[10]	L1 [v]	vielleicht. (• •) Also, das ste/ wenn Junge ist, dei dein und appel Apfel, (•)	
[11]	L1 [v]	Junge	(3s) dann lat das / kennst du das, wenn man irgendwie / (2s)
	L2 [v]	Mhm.	Mhm.
[12]	L1 [v]	(•••) Weiß nicht, also ich würde sagen, das heißt irgendwieäh Junge las dei / dein	
	L1 [nv]		<i>lachen</i>
[13]	L1 [v]	dein Ap / deinee Äpfel Junge lass deine Äpfel (• • •) stahn.	
	L2 [v]	(•) las / las dein	stan standen
[14]	L1 [v]	Stehen, ja.	Junge lass deine ne Äpfel stehen. (2s) Ja?
	L2 [v]	stehen stahn. (• • •) äha.	lass deine Äpfel stehen
[15]	L1 [v]	Jaa?	Ja, also ich weiß es auch nicht. Okey. Ähm. So.
	L1 [nv]	<i>lachend</i>	
	L2 [v]	Jaa? Vielleicht.	Ja? Ähm?
	L1 [nv]	<i>lachen</i>	

durch Erweiterung des Kontextes und proaktive Inputmodellierung. Es zeigt sich in unseren Daten zudem ein Verfahren, das Vandergrift (2016: 187) als *forward inference* bezeichnet. Dabei werden durch Indizierungen, die vorwärtsgerichtet projektiv wirken, passende strukturierende Hinweise gegeben, wodurch gleichzeitig eine von der L2-Sprecherin einfach zu füllende Lücke generiert wird, die lernförderlich ist. Dadurch bekommt die Sequenz einen aufeinander aufbauenden, synchronisierenden und lernförderlichen Charakter.

Ein wichtiger Mehrwert des virtuellen Settings liegt im empraktischen Handeln, denn zur Überprüfung einer möglichen Lösung sind die TP gezwungen, mehrere Quellen zu nutzen, sich in den benachbarten Stationen synchronisiert fortzubewegen und mehrere interpretative Hinweise zu koordinieren, was nicht nur interaktional und aktional eine wichtige *joint action* und kollaborative Exploration darstellt, sondern auch kognitiv als Phase der Reorientierung und Strukturierung zur kognitiven Dynamisierung und Entspannung sowie zur prägnanten Fokussierung dient. Dies scheint für die DaF-Lernende wichtig bei der Verankerung neuer inferenzieller Strategien. So bewegen sich beide Tandempartnerinnen im vorliegenden Beispiel (da lokale Informationen nicht ausreichen und eine kontextuelle Interpretation notwendig ist) explorativ durch den virtuellen Raum zur möglich hilfreichen Quelle aus der vorherigen Station. Insgesamt



wird die L2-Lernende durch diese kollaborative Aktivität angeregt, nicht nur form-basiert und anhand lokaler Hinweise zu interpretieren, sondern die kontextuelle, inhaltliche Passung zu beachten und interaktive Nachvollziehbarkeit durch Begründungen herzustellen.

#### BEISPIEL 4:

[25]

L1 [v] Würde es für dich Sinn ergeben, also weißt du, was damit gesagt werden soll?

[26]

L2 [v] (•••) Ähm, aber ich verstehe so Huckup nicht so (••) so so so gut. Äh, ich

[27]

L1 [v] Jaa, das war glaube ich / warte mal ganz kurz. Ich

L2 [v] glaube es ist ein Denkmal, oder?

[28]

L1 [v] Warte mal ganz kurz, ich guck lieber noch mal nach. (3 s) Ach so, äh (2 s) Ja, Denkmal an

[29]

L1 [v] der Schuhstraße.

Genau ja, du hast Recht.

L2 [v] (2 s) Also Hochdeutsch Huck auf, ja.

nonverbal

*Beide Avatare springen zu dem Einstiegstext.*

Die kollaborative, empraktische Sequenz dient somit zur Kontextualisierung und Überprüfung der Gültigkeit und ermöglicht die Verknüpfung des lokalen Deutungskontexts mit der Semantik einzelner Elemente und dem ganzen Script. Empraktisch fundiert, durch die synchronisierende Bewegung im virtuellen Raum, ist die neue mentale Brücke im inferenziellen System der Lernenden prägnant angebahnt. Die sehr abstrakte und im klassischen Unterrichtskontext schwer trainierbare Strategie der Verknüpfung lokaler, induktiver und globaler, script-basierter Kohärenzstrategien wird im virtuellen Raum in einem kollaborativen, explorativen und expressiven Akt mit einem zugleich wichtigen integrativ-interaktiven Wert sehr prägnant gefördert. Neben diesen kurzfristigen Erwerbsgewinnen sollten die Auswirkungen der Lernumgebung aus longitudinaler Sicht auf sprachliche Reflexivität und selbstregulative Prozesse untersucht werden, wie auch das Transferpotenzial auf die reale Kommunikation.

#### 4.3. Ko-Konstruktionen als interkulturelle Lernressource

Da die Tandempartnerinnen ihre Handlungsräume flexibel und selbstbestimmt ko-konstruieren, kommt es häufiger vor, dass nach *Gaming*-Phasen mit ausgeprägter Koordinationsarbeit der Raum um narrative, phatische Sequenzen oder interkulturell vergleichende Nebensequenzen expandiert wird. Diese sorgen einerseits für kognitive und interaktive Entspannung und können andererseits als ein Beweis zunehmender interaktiver Nähe und integrativen Handelns gelten. So werden die Ähnlichkeiten zwischen der Sage der standarddeutschen Zielsprache und einer spontan eingebrachten Geschichte aus der Sprache(n) der Lernerin thematisiert. So entwickelt sich aus einer sprachreflexiven eine kulturelle Episode, in der die TP nicht nur als Kulturvermittler auftreten, sondern in der Deutungsarbeit gleichzeitig eine gewisse Distanz und

Neuperspektivierung zum Eigenen erfahren. Aus einem selbstzentrierten Verhalten entwickelt sich eine synchronisierende, um intersubjektive Verständlichkeit bemühte Aktivität: Die TP paraphrasieren und handeln insgesamt verständnissichernd und multiperspektivischvergleichend. Das virtuelle Setting fördert sie, mehr zu explizieren, sich zu synchronisieren und somit das vermeintlich Bekannte dem Adressaten (deutend) anzupassen und anschließend ko-konstruktiv an der Wissens- und Differenz in der Argumentation zu arbeiten. Die kollaborative Anpassung ermöglicht so die Arbeit an potenziellen Widersprüchen und die Neuperspektivierung des Eigenen. Bei der Arbeit an der Brücken-Herstellung zwischen kultur- und sprachspezifischen Aspekten aktivieren die TP damit zusammenhängende Deutungsmuster und bilden konzeptionelle Brücken (vgl. die diskursive Interkultur, Koole & ten Thije 1994) und lernen diese sprachlich angemessen zu formulieren. Es zeigt sich eine sukzessiv zunehmende Berücksichtigung der Perspektive des Anderen in den Prozessen der Deutung, Verständnis- und Bedeutungsaushandlung sowie in der interkulturellen Anpassung durch ko-konstruierende Arbeit an der Divergenz auf. Diese hat nicht nur eine wichtige sozialisierend-integrative Funktion, sondern ermöglicht eine Flexibilisierung und Diversifizierung der Deutungsmuster (Altmayer 2016).

Im folgenden Beispiel wird dies exemplarisch gezeigt: Nach der Darstellung einer als partiell äquivalent empfundenen Legende der eigenen Kultur wird ein Sprichwort als kondensierte Pointe erwähnt (Fl. 19, 20). Die L2-Tandempartnerin befürchtet allerdings, nicht verstanden zu werden (vgl. Fl. 20 *Kannst du nicht ganz verständlich? (lachen)*) bzw. den narrativen Höhepunkt interaktiv nicht erfolgreich gestaltet zu haben. L1 unterstützt den narrativen Höhepunkt durch symmetrisches Fortführen, bietet verständnissichernd ein mögliches Äquivalent der Zielsprache an und initiiert durch den Versuch einer Annäherung (vgl. was *Ähnliches*) eine kollaborative Arbeit an der konzeptionellen Brückenbildung (Fl. 22). Auf diese reagiert L2 zuerst mit einer expressiven Bestätigung (*Ah, ja!*) und bewegt sich danach zur vertieften Arbeit an dem konzeptionellen „*Matching*“.

### BEISPIEL 5:

[18]

L1 [v] Ja.

L2 [v] hört das anderen, das klingen. Deshalb haben wir also einen so solche Sprichwort oder so,

[19]

L2 [v] ähm zum Beschreiben, also man macht das ähh / also ähm man muss über etwas schlechtes Handlung

[20]

L2 [v] verstecken, aber trotzdem es wird von anderen bemerkt. Kannst du nicht ganz verständlich? (lachen)

[21]

L1 [v] Ah, also so was wie äh/ Was ähnliches wie äh "Lügen haben kurze Beine". Also das man /was

[22]

L1 [v] Ähnliches?

Ja.

Ja.

L2 [v] Ah. Ja! Aber vielleicht ja, aber es ist ein böse Handlung oder so. Und er handelt.

[23]

L1 [v] Wenn man etwas Böses macht, dass man dass das aufgedeckt wird, auch wenn man versucht es zu

L2 [v] Ja.

Ja.

[24]

L1 [v] es zu verstecken. Cool. Ja. Wie sagt man das auf Chinesisch? Es klingt voll gut.

L2 [v] Ja. Cool.

Ähh. Huǎngyán tuǐ duǎn. Lachen (3s)



L2 erkennt durch das Verständnis der Interaktionspartnerin eine konzeptionelle Differenz und wird zur vertieften Reflexion, zur Reoperspektivierung, Bedeutungsadjustierung und Ausdifferenzierung angeregt. Die Feinadjustierung führt zur Isolierung zweier Aspekte, die für L2 konzeptionell nicht äquivalent sind bzw. sie entwickelt somit eine neue Wahrnehmung des Eigenen. Die epistemische Relativierung (*vielleicht, oder so*) öffnet diese dissonanten Aspekte interaktiv für weitere Ko-Konstruktionsakte, wonach die L1-Sprecherin kollaborativ ausbauend fortführt und dabei eine sprachliche Neurelationierung vornimmt. Nun wird die bekannte Bedeutung für das deutsche Äquivalent anschließbarer formuliert bzw. kulturell und sprachlich unterscheidenden Metaphern und Scripts angenähert. Dies hat einen sprachlich modellierenden Charakter, denn es nimmt die Äußerung von L2 und die vom L2-TP selbst pointiert hervorgehobenen differenten Aspekte auf. Der ko-konstruktive Charakter bei der deutenden Annäherung und Synchronisierung im Verstehen (vgl. *ja, cool*) trägt zur Maximierung der interaktiven Nähe bei und ermöglicht neue konzeptionelle Verknüpfungen fremdsprachlicher Strukturen und tieferliegender Konzepte. Motiviert durch diese ko-konstruktive Leistungen initiieren TP weitere sprachreflexive und kulturvergleichende Episoden.

Dies zeigt eindeutig, dass das Tandem auf dem Austausch, der Offenheit, Vielfalt und dem interkulturellen Interesse beruht. L2 wird in ihrem Deutungshorizont an keiner Stelle unterminiert und zur Anpassung an die Zielsprache gezwungen, sondern es wird ko-konstruktiv am gemeinsamen Verständnis gearbeitet. Approximative Angebote, *Grounding* durch Hypothesentesten (*vorwärtsgerichtete Inferenz*, vgl. Vandergrift 2016: 187) und symmetrisches Fortführen treiben L2 zur Selbstanpassung, zur Hörerorientierung und zum Perspektivenwechsel voran. L2 lernt nicht nur konvergent zu handeln, sondern auch Widersprüche, Divergenzen zu erkennen und diese als Ressource für die kollaborative Bedeutungsaushandlung zu nutzen. Diese Lernerfahrungen tragen zur Entwicklung interaktiver Routinen im Diskursverlauf bei und fördern Kompetenzen zur ko-konstruktiven, interkulturell angepassten Bedeutungsaushandlung. Die Erfahrungen der kooperativen sprachlich-konzeptionellen Brückenbildung und interaktiven Anpassung werden durch das synchronisierende, multimodale VR-Setting prägnanter wahrgenommen, was auch die Befragungen der TP bestätigt haben (Kapitel 4.1.): etwa wenn die TP über die tiefe, nahe, intensive Interaktion sowie über das Gefühl der Internalisierung der Perspektive des Anderen berichten, „direkter die andere Denkweise zu kennen und [sich] anzueignen“.

## 5. ZUSAMMENFASSUNG UND AUSBLICK

Das hier untersuchte Setting der virtuellen Lernwelt in der *Hololingo!*-App hat sich als Grundlage einer kommunikativen, unterhaltsam-spielerischen, empraktisch-handlungs-orientierten und fokussierten Tandemarbeit gezeigt, welche in mehrerer Hinsicht lernförderlich ist. Sie ermöglicht (im Vergleich zu oft überfordernden, unfokussierten *Hangout*-Tandems in herkömmlichen SVR-Apps) eine intensivere, symmetrische,

abwechslungsreiche, zielgerichtete und selbstregulative Interaktion mit mehr Anbindungspotenzial an die institutionellen Lernprozesse. Die Tandemarbeit in der *Hololingo!*-App wird durch die Nutzerinnen – trotz einiger Optimierungsvorschläge – als große Bereicherung erlebt. Im Feedback nennen sie viele lernförderliche Aspekte, unter anderem die intensivere Lernerfahrung in allen wichtigen (interaktiven, emotionalen und aktionalen) Dimensionen des Lernprozesses. Sie benennen aber genauso das neuartige empraktische Handeln in der kognitiv neu zu bearbeitenden virtuellen Realität bzgl. der Synchronisierungen mit dem Interaktionspartner in Wahrnehmung, Bewegung und Aktion, die für eine prägnantere und tiefere Lernerfahrung sorgen. Wir führen die beschriebenen Wirkungen u. a. auf die explorative, narrativ-gesteuerte, selbstregulative Einbindung, den Lernabenteuer-Charakter und das Angebot abwechslungsreicher kollaborativ-kommunikativer Aufgabenstellungen sowie zur Sprachreflexion im Gaming-Modus zurück. Ein weiterer wichtiger Faktor dabei ist die integrierte Phase zur Vorbereitung der Selbstregulation. Diese gliedert sich einerseits in die vorgelagerte Aushandlung der Tandemrollen zu Beginn des Lernabenteuers (anhand eines Videoinputs) und andererseits in die integrierten und nachgelagerten Reflexionsphasen, welche die TP für die Funktion der Tandemarbeit, ihre jeweiligen Akteursrollen und die damit verbundene Selbstverantwortung ihres eigenen Lernprozesses sensibilisierte. Der Erfolg DGBLL-basierten Tandemlernens in der VR wird künftig vor allem von der gezielten Entwicklung kommunikations- und erwerbsförderlicher Aufgabendesigns abhängen, die sehr spezifisch einzelne sprachliche Phänomene evozieren, aber auch holistisch kommunikativ einbinden und komplexe Anforderungen an Tandems stellen. Nur so kann einerseits eine festere curriculare Anbindung an den DaF-Unterricht gelingen, etwa durch Lernwelten, die sich an den Niveaustufen und curricularen Inhalten des GER orientieren, bei gleichzeitiger angewandter, holistischer Sprachförderung. Obwohl erst eine größere Datengrundlage ein valideres Bild ergeben kann, wurde anhand der qualitativen Interaktionsanalyse deutlich, wie unterschiedlich einzelne Aufgabentypen die Interaktionen vorantreiben. Im SVR-Tandem erworbene diskursive Kompetenzen lassen sich vermutlich weitgehend auf analoge Kommunikationskontexte transferieren, auch wenn hier künftige Forschung intensiver anknüpfen muss.

Die Interaktionsanalyse aus einer spracherwerbsförderlichen Perspektive zeigt, dass die Tandems intensive, mehrdimensionale und lernförderliche Aushandlungsprozesse betrieben, bei denen die kollaborative Selbststeuerung des Lernprozesses durch viel Eigeninitiative von L2-TP und zurückgenommenes, gezieltes supportives Sprachhandeln von L1 zu einer holistischen Lernerfahrung auf vielen sprachlich-kommunikativen Ebenen führte und insbesondere für interkulturelle Lernprozesse förderlich zu sein schien. L1 ließ hier L2 viel Raum zur Selbststeuerung in Bezug auf Inhalte und Schwierigkeitsgrad. Die TP handeln in vielen unterschiedlichen Handlungsräumen, selbstregulativ, ko-konstruktiv, sprachreflexiv, kreativ, in der Bedeutungsaushandlung in mehreren Hinsichten brückenschlagend und argumentativ-lösungsorientiert. Neben gezielter Förderung kommunikativer, interkultureller und bestimmter lexikalischer oder syntaktischer Kompetenzen zeigt sich das Setting besonders produktiv zum Training abstrakter inferenzieller Kohärenzherstellungsstrategien durch kooperative, sprachreflexive Entschlüsselungs- und Deutungsaufgaben, für die neben Texten auch multimediale Objekte herangezogen wurden. Auch wurde dazu empraktisches Handeln mit kognitiv herausfordernden Aufgaben kombiniert.

Die besondere Lernqualität dieser Lernumgebungen scheint durch das *Embodiment*, das empraktische Handeln und die Herstellung interaktiver und virtuell-proximaler Nähe

durch Gestik, Positur und Emojis an einem (wenn auch virtuellen) körperlich erfahrbaren Ort bedingt zu sein. Durch die technisch bedingte Reduktion von Mimik, die in Bezug auf den Transfer einen Nachteil darstellen kann, beobachteten wir eine stärkere Fokussierung auf sprachliche und metasprachlich-inhaltliche Aspekte (Prosodie, Artikulation) und Gestik, die der Tandemarbeit möglicherweise sogar förderlich zugutekommt. Genauso werden zur Kompensation des fehlenden mimischen *expression-displays* viele neue, expressiv und emotional sogar intensivere Formen genutzt, die zusammen mit dem intensiven Synchronisierungseffekt für eine intensive affektive Involvierung und somit emotional-basiertes Lernen sorgen. Auch konnten Phänomene an der Schnittstelle von Interaktion und Kognition beobachtet werden, u.a. multifunktionale *moves-of-selfregulation*. Dieser Aspekt wäre in künftigen Untersuchungen fokussierter zu prüfen.

Nächstes Ziel der Pilotstudie *Hololingo!* ist es nun, weitere dynamisch-immersive Handlungsräume umzusetzen (als Gamifizierung von Lexik, Grammatik, Pragmatik) sowie mit weitere Episoden zur curricularen Anbindung zu entwickeln. Nachfolgeuntersuchungen sollten zudem die erkannten Besonderheiten des Lernens im SVR-Tandem noch differenzierter und anhand einer größeren Datenlage fokussieren sowie mit Hilfe triangulativer Verfahren erweitern, beispielsweise in der Kombination einer Video- und Eye-Tracking-Studie mit Interviews und Lautdenkmethode. Auch sollen evaluative Pre-/ Posttests ergänzt werden, um den nach dem nächsten Entwicklungsschritt reiferen Demonstrator bzgl. kurz-, mittel- und langfristige Lernprozesse genauer zu erfassen. Auch für Lehrkräfte (in Ausbildung) kann es hilfreich sein, durch eigene SVR-Nutzung die Perspektive der Lernenden besser zu verstehen und sich somit stärker für die Rollen und tandemförderlichen Sprachhandlungen in der ko-konstruktiven Unterstützung und Inputmodellierung zu sensibilisieren, eigene digitale Kompetenzen zu erweitern und ein neues Bewusstsein für die Nutzung virtueller Ressourcen zu entwickeln. Neben der Auswahl durch Lernende und Lehrende von Inhalten in DGBLL-SVR-Apps muss auch eine stärkere Verknüpfung (Vorbereitung, Auswahl, Transfer) mit analogen Lernkontexten bisheriger DaF-Praxis erfolgen.

### Bibliographie

1. Ahlers, Timo / Bumann, Cassandra / Kölle, Ralph / Lazović, Milica (2021): "Hololingo! - A Game-Based Social Virtual Reality Application for Foreign Language Tandem Learning!". In: Kienle, Andrea; Harrer, Andreas; Haake, Jörg M.; Lingnau, Andreas (Hrsg.): Die 19. Fachtagung Bildungstechnologien der Gesellschaft für Informatik, Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021, 37-48.
2. Ahlers, Timo / Siegert, Gregor (2019): „Sprachimmersion im Wohnzimmer: Interaktion, Grounding und Embodiment im DaF-Erwerb mittels Social-VR“. In: Philipp, Hannes; Weber, Bernadette; Wellner, Johann (Hrsg.): Kosovarisch-rumänische Begegnung. Beiträge zur deutschen Sprache in und aus Südosteuropa. Regensburg: Universität Regensburg, 94-117 (Forschungen zur deutschen Sprache in Mittel-, Ost- und Südosteuropa FzDiMOS 8).
3. Ahlers, Timo; Lazović, Milica; Schweiger, Kathrin; Senkbeil, Karsten (2020): "Tandemlernen in Social-Virtual-Reality: Immersiv-spielebasierter DaF-

- Erwerb von mündlichen Sprachkompetenzen". In: Zeitschrift für Interkulturellen Fremdsprachenunterricht 25 (2), 237–269.
4. *Altmayer, Claus et al.* (2016) (Hrsg.): *Mitreden: diskursive Landeskunde für Deutsch als Fremd- und Zweitsprache*. Stuttgart, Ernst Klett Sprachen.
  5. *Bechtel, Mark* (2010): „Sprachentandems“. In: Weidemann, Arne; Straub, Jürgen; Nothnagel, Steffi (Hrsg.): *Wie lehrt man interkulturelle Kompetenz. Theorien, Methoden und Praxis in der Hochschulausbildung. Ein Handbuch*. Bielefeld: Transcript, 285–300.
  6. *Bertrand, Philippe; Guegan, Jérôme; Robieux, Léonore; McCall, Cade A.; Zenasni, Franck* (2018): *Learning Empathy Through Virtual Reality: Multiple Strategies for Training Empathy-Related Abilities Using Body Ownership Illusions in Embodied Virtual Reality*. In: *Frontiers in Robotics and AI*, 1–18.
  7. *Biebighäuser, Katrin* (2013): „Fremdsprachenlernen in virtuellen Welten. Aufgabengestaltung in komplexen multimodalen Lernumgebungen“. In: *Fremdsprachen lehren und lernen (FLUL)* 42 (2), 55–70.
  8. *Biebighäuser, Katrin* (2014): *Fremdsprachenlernen in virtuellen Welten. Empirische Untersuchung eines Begegnungsprojekts zum interkulturellen Lernen*. Tübingen: Narr.
  9. *Biebighäuser, Katrin; Feick, Diana* (Hrsg.) (2020): *Digitale Medien in Deutsch als Fremd- und Zweitsprache*. Berlin: Erich Schmidt Verlag.
  10. *Blume, Carolyn; Schmidt, Torben; Schmidt, Inke* (2017): „An Imperfect Union? Enacting an Analytic and Evaluative Framework for Digital Games for Language Learning“. In: *Zeitschrift für Fremdsprachenforschung* 28 (2), 209–231.
  11. *Böcker, Jessica* (2013): Reflexionen über das Fremdsprachenlernen im Tandem am Beispiel des Bochumer Tandem Logbuchs. In: Hahn, Natalia, Reinecke, Katja (Hrsg.): *Erfahrungen mit Sprachlerntandems. Beratung, Begleitung und Reflexion. Beiträge der Tandem-Tagung an der Pädagogischen Hochschule Freiburg/Br. 2012*. Freiburg: Pädagogische Hochschule, 7–30.
  12. *Böcker, Jessica; Kleppin, Karin* (2017): Das Lernerlogbuch in den binationalen Kursen des Deutsch-Französischen Jugendwerks (DFJW). In: Böcker, Jessica; Ciekanski, Maud; Cravageot, Marie; Kleppin, Karin; Lipp, Kai-U. (Hrsg.) (2017): *Kompetenzentwicklung durch das Lernen im Tandem. Akteure, Ressourcen, Ausbildung. Eine deutsch-französische Studie*. Paris; Berlin: Deutsch-Französisches Jugendwerk (= Arbeitstexte 29), 51–92.
  13. *Christ, Herbert* (1996) Das nahe Fremde und das ferne Fremde im fremdsprachlichen Literaturunterricht. In: Lothar Bredella; Herbert Christ (Hrsg.) *Begegnungen mit dem Fremden*. Gießen: Verlag der Ferber'schen Universitätsbuchhandlung, 89–107.
  14. *Ehlich, Konrad* (2007): *Sprache und sprachliches Handeln*. Bd. 1 – 3. Berlin: De Gruyter.
  15. *Funk, Hermann; Gerlach, Manja; Spaniel-Weise, Dorothea* (Hrsg.) (2017): *Handbook for Foreign Language Learning in Online Tandems and Educational Settings*. Frankfurt a. M.: Peter Lang (Foreign Language Teaching in Europe 15).
  16. *Hartfill, Judith; Gabel, Jenny; Neves-Coelho, Daniel; Vogel, Daniel; Räthel, Fabian; Tiede, Simon; Ariza, Oscar; Steinicke, Frank* (2020): „Word saber: an effective and fun VR vocabulary learning game“. In: Preim, Bernhard; Nürnberger, Andreas; Hansen, Christian (Hrsg.): *Tagungsband Mensch und Computer 2020 (MuC'20)*. New York: Association for Computing Machinery, 145–154.
  17. *Hung, Hsiu-Ting; Yang, Jie C.; Hwang, Gwo-Jen; Chu, Hui-Chun; Wang, Chun-Chieh* (2018): „A scoping review of research on digital game-based language learning“. In: *Computers & Education* 126, 89–104.
  18. *Koole, Tom / Thije, Jan D ten* (1994): *The Construction of Intercultural Discourse. Team discussions of educational advisers* (Utrecht: diss.). Amsterdam / Atlanta.
  19. *Mystakidis, Stylianos* (2019): *Motivation enhanced deep and meaningful learning with social virtual reality*. University of Jyväskylä (Dissertation). Online: <https://jyx.jyu.fi/handle/123456789/66667> [06.03.2021].
  20. *Redder, Angelika / Rehbein, Jochen* (1999) (Hrsg.): *Grammatik und mentale Prozesse*. Tübingen: Stauffenburg-Verlag.



21. *Vandergrift, Ilona* (2016): *Second-language Discourse in the Digital World. Linguistic and social practices in and beyond the networked classroom.* Amsterdam/Philadelphia: John Benjamins.
22. *Zender, Raphael; Knoth, Alexander H.; Fischer, Martin H.; Lucke, Ulrike* (2019): "Potentials of Virtual Reality as an Instrument for Research and Education". In: *i-com* 18 (1), 3–15.
23. *Zhao, Shanyang*. 2003. "Being there" and the Role of Presence Technology. In Giuseppe Riva, Fabrizio Davide & Wijnand A. IJsselstein (eds.), *Being There: Concepts, effects and measurement of user presence in synthetic environments*, 137–146. Amsterdam: IOS Press.

### **DAF IM TANDEMLERNEN MIT DER HOLOLINGO!-APP. EINE ANALYSE DER TANDEMKOMMUNIKATION IN GAME-BASED SOCIAL VIRTUAL REALITY<sup>6</sup>**

*Lazovic Milica, Dr. phil., Philipps-Universität Marburg,  
Deutschland, milica.lazovic@uni-marburg.de*

*Ahlers Timo, Dr. phil., Universität Potsdam, Deutschland,  
timo.ahlers@uni-potsdam.de*

Tandemlernen in Social Virtual Reality erweitert die Praxis des Fremdspracherwerbs um viele innovative Aspekte. Beginnend mit einem allgemeinen Überblick zu SVR-Praktiken und ihren Lernpotenzialen möchten wir die von unserer Forschungsgruppe entwickelte Hololingo!-App vorstellen und anschließend einen Einblick in DaF-Tandeminteraktionen geben. Dabei fokussieren wir aus einer interaktionsanalytischen Perspektive einige erwerbsrelevante Handlungsaspekte in der VR-Umgebung, bedingt einerseits durch das empirische Handeln und andererseits in einer sprachreflexiven Episode des Spiels, in der Tandempartner kollaborativ an der Deutung einer Inschrift sowie kulturreflexiv agieren. Aus einem quantitativen Vergleich mit *Hangouttandems* (in *AltspaceVR*) bzgl. der Diskurspartizipation sowie anhand der Auswertung von Nutzerfeedback leiten wir weitere Potenziale Gaming-basierter VR-Anwendungen für die künftige Tandempraxis ab.

**• DaF • Tandem lernen • Hololingo! • Social Virtual Reality.**

<sup>6</sup> Das Projekt wurde gefördert durch die Universität Hildesheim (Anschubfinanzierung) und den Stifterverband [Programm: Wirkung hoch 100, Projekt: H110 5114 5132 36419, Förderzeitraum: 12/2020-03/2022]. Wir danken allen Proband:innen und involvierten studentischen Hilfskräften (C. Bumann, M. Goecht, A. Schwanke, K. Köller und F. Breker), dem Kollegen Ralph Kölle sowie den weiteren Mitgliedern der Forschungsgruppe *Hololingo* (<https://hololingo.blog.uni-hildesheim.de>) für ihre Beteiligung bei der Konzeption und Entwicklung der *Hololingo!*-App.

# Прощайте, Алексей Михайлович!..



В это трудно поверить, ещё труднее примириться: ушёл из жизни АЛЕКСЕЙ МИХАЙЛОВИЧ КУШНИР — педагог, учёный, журналист, главный редактор журнала «Народное образование» и директор одноимённого Издательского дома. Он был сгустком энергии, новых идей, постоянного движения. Многое оставил на земле. И в детях, и в трудах.

Формат журнала, его отношение к актуальным проблемам педагогики и образовательной политики, его верность педагогике дела, наследию Антона Макаренко, такие проекты, как спецвыпуски, посвящённые юбилеям А.С. Пушкина, Великой Отечественной войны и Отечественной войны 1812 года, истории журнала, — всё это Алексей Михайлович Кушнир.

Нельзя не вспомнить, как находил и приветствовал новых авторов разных поколений, для многих из которых «Народное образование» стало настоящей школой.

Назову Виктора Трофимовича Чумакова, Юрия Крупнова, Анатолия Вифлеемского, себя, Арсения Замостьянова, покойного нашего друга Дмитрия Григорьева. И долго можно продолжать этот список авторов кушнирского «Народного образования». Всех нас Алексей Михайлович понял, собрал, ради нас с лёгкостью рисковал, доверял нам. И получились, смею думать, неплохие номера, выпуски, вошедшие в историю просветительской печати. Вместе мы создали историю старейшего журнала



России, созданного ещё во времена Александра I. У какого ещё журнала есть столь капитальный летописный свод?

Только на одно ему вечно не хватало времени — на публикацию собственных книг, на защиту докторской диссертации, которую он — как учёный — давно перерос. Но мы уверены, что к его статьям и книгам ещё долго будут возвращаться исследователи, студенты, школьники.

Он всегда был смел — и в поступках, и в теоретических построениях. Отстаивал свою правду до последнего дня. Это ярко проявилось в ежегодных Макаренковских чтениях, да и вообще — в каждой статье Алексея Кушнера.

Вечная память выдающемуся педагогу нашего времени, замечательному редактору, нашедшему и сплотившему десятки ярких авторов, настоящих профессионалов. А его новаторские идеи принадлежат будущему. Мы обязательно будем подробнее рассказывать о них.

В последние годы ему было трудно. Таких людей уносят не болезни, да и было ему всего лишь шестьдесят два. Он был сильным человеком, а таких не жалеют. И подчас взваливают на их плечи неподъёмную ношу. А Алексей Михайлович был честным и ранимым человеком, хотя это не бросалось в глаза. Он по-мужски скрывал свои несчастья, свои тяжёлые мысли. И до конца оставался победителем, замечательным редактором и педагогом, полным новых идей. Он всё равно останется живым для нас. Мы будем советоваться и спорить с ним, как прежде. Не может такой человек уйти.

Смириться с этой утратой невозможно.

Вечная память!

*Арсений Замостьянов, коллеги,  
редакция журнала «Народное образование»*

# Юрий Николаевич Ромашкин

(30.X.1953–25.X.2021)



25 октября 2021 года после непродолжительной тяжелой болезни скоропостижно скончался член редколлегии журнала «Речевые технологии», кандидат технических наук, старший научный сотрудник Юрий Николаевич Ромашкин.

Научная и производственная деятельность Юрия Николаевича началась в 1980 году после окончания с отличием Московского инженерно-физического института. С начала своей деятельности он проявил себя активным, вдумчивым исследователем, стремившимся к освоению и внедрению современных речевых технологий. Его глубокие знания и эрудиция стали основой его известности и высокого авторитета в среде ученых и специалистов в области речевой акустики.

Порядочность, отзывчивость, честность и принципиальность, большой опыт научных исследований, знания в смежных областях, профессионализм и ответственность за порученное дело — качества, которые отличали Юрия Николаевича и вызывали искреннее уважение всех его коллег.

Юрий Николаевич был жизнерадостным, душевным человеком, обладал положительной энергетикой, умел общаться с людьми на различные темы. Окружающие и близкие ценили его прекрасные качества и внутреннюю душевную теплоту. Память о нем будет светлой и навсегда сохранится в наших сердцах.

## Content

*Pantiukhin D.V.*

**Neural networks for speech synthesis of voice assistants and singing machines . . . . . 3**

*Kharchenko S.S., Kryuchkov I.A., Meshcheryakov R.V.*

**Determining the pitch frequency of a speech signal based on the multidimensional synchronization index . . . . . 17**

*Borisov V.V., Kharlamov A.A.*

**Intellectual analysis and modeling of the development of situations based on a combination of neural network word processing, methods of fuzzy cognitive analysis and dynamic clustering . . . . . 29**

*Seyedmilad Ranaei Siadat.*

**Emotion recognition in Persian speech using a one-dimensional (1D) neural network . . . . . 44**

*Pilgun M.A.*

**Cognitive research in real time: neural network approach . . . . . 57**

*Koreneva O., Limbach K., Shtender A., Gore M.*

**Neurodidactics and the teaching of foreign languages and translation in the media space – on the example of the German podcast of the University of Spain . . . . . 71**

*Ana Medina Reguera.*

**Speech technologies for the development of communication in children with a lack of speech . . . . . 84**

*Isabelle Gallego Gallardo.*

**The concept of "sustainable tourism" in the social, economic, environmental and linguistic spheres . . . . . 100**

*Milica Lazovic, Timo Ahlers.*

**Teaching German as a foreign language in tandem with Hololingo!: An analysis of tandem communication in social virtual reality gaming . . . . . 113**

**In memory of A.M. Kushnir . . . . . 135**

**In memory of Yu.N. Romashkin . . . . . 137**

## Система анализа поведения людей на соответствие установленным правилам и регламентам SecurOS SBA (Semantic Behavior Analyzer)

Система собирает, обрабатывает и анализирует видеоданные, полученные от стереокамер; выдает оповещения при детекции действий или бездействии, нарушающих регламенты, а также опасных ситуаций; преобразует результаты анализа в интерактивные карточки событий и табличные отчеты, а также отображает статистику инцидентов в виде дашбордов (графиков и диаграмм).

Интерактивные карточки событий дают возможность специалистам служб безопасности и оперативного контроля незамедлительно реагировать на происшествия. Отчеты и дашборды предназначены для управляющих сотрудников и позволяют делать выводы о ситуации на конкретных участках или на предприятии в целом.

### Архитектура решения



### Ключевые преимущества решения

- SecurOS SBA дает возможность учитывать поведенческие характеристики людей с привязкой к элементам сцены, а также причинно-следственные связи во взаимодействиях между людьми или человеком и объектом (-ами). Таким образом, глубина и точность анализа при использовании технологий SecurOS SBA значительно превышает возможности традиционной видеоаналитики, что существенно расширяет круг применений решения.
- Лингвистический процессор преобразует текст регламентов в работающие правила на встроенной экспертной системе, т.е. без необходимости создания специализированного программного обеспечения под каждый проект.





**Индекс: 62203**  
**ISSN 2305-8129**