

DOI 10.23683/2415-8852-2018-2-51-64

УДК 316

АССИСТЕНТ И ЕГО ДОЛЖНИК: ОБ ИСКУССТВЕННЫХ ГОЛОСАХ



Евгения Валерьевна Самостиенко

старший преподаватель Института филологии и журналистики ННГУ им. Н.И. Лобачевского, Нижний Новгород
e-mail: eugeniasuslova@gmail.com

Аннотация. Статья посвящена исследованию искусственных голосов и голосовых агентов в цифровую эпоху, для которой характерны такие особенности, как дискретизация информации, перевод качественного в количественное, распыленность, распределенность (замена контактного дистантным), замена линейности нелинейностью. Парадоксальность цифрового голоса возникает по причине того, что голос, с одной стороны, всегда указывает на субъекта (говорит всегда кто-то), а с другой стороны, говорит потому, что несет определенное желание, иначе сам акт говорения не возник бы. В отличие от них искусственные голоса предлагают особую модель агентивности и распределения желания, а также придают речи особый статус – статус внутренней речи. В статье мы рассмотрим инструменты для создания компьютеризированных голосов (Govorilka), инструменты синтеза и редактирования речи (Adobe VoCo, WaveNet), голосовые интерфейсы в компьютерных играх, очертим контур функционирования голосовых агентов, а также рассмотрим голос как биометрический феномен, используемый для идентификации субъекта.

Ключевые слова: голосовой агент, искусственный голос, синтез речи, голосовой ассистент, желание.

В этой статье речь пойдет о голосовых агентах и искусственных голосах, о том, какое место они занимают среди нас и как влияют на переживание контакта и практики использования языка. Меня интересует вопрос о том, как искусственные голоса вплетаются в тонкую ткань социального. Думать, что речь может прояснить причины действия, было бы наивно – скорее, она всегда проходит по невидимым силовым линиям. Подчиняется ли искусственная речь их логике и какова конфигурация этих линий?

Проводники желания

Если человек молчит, это значит, что в данный момент у него нет никакой потребности, никакого желания. Голос всегда обращен к другому и выступает проводником желания. Если это так, то что делать с голосами машин, ведь они не могут желать? Они не желают, но выражают желание. Можем ли мы поставить вопрос о том, чье оно?

В попытке ответить на этот вопрос мы сталкиваемся со специфической структурой желания. Принципиальное отличие голосовых ассистентов, с которыми мы сегодня имеем дело, вроде Siri или Алисы, заключается в том, что они никогда не являются источниками голоса в том смысле, в котором не являются источниками желания, но в то же время не формируют никакой автономии вроде объекта производства, как это бывает, скажем, с программами для создания ви-

зуального контента. Наоборот, они всегда откликаются на желание другого, в данном случае – на нехватку, на желание знать, которое позволит совершить субъекту то или иное действие (выбрать кафе, найти номер в телефонной книге и т. д.). Но помимо этого искусственные голоса также указывают на язык как технологию доступа к себе, на особые модели внимания и распознавания и производят пересборку внутренних структур социального.

Голосовые машины: от жутких кукол к цифровым ассистентам

Парадоксальным образом голос аннулирует акт копирования. Там, где В. Беньямин видел противопоставление оригинала и копии, в случае голоса речь идет о повторе. Повторение, своего рода эхо, указывает на смещение с содержания звучащего на сам акт звучания. Голос находится ровно посередине между содержанием речи и полем, внеположным субъекту. «Акустическое зеркало» (К. Сильверман) выступает своеобразной заменой рефлексии – круг *голос – ухо* моделирует пространство и «ощупывает» потенциальное, одновременно указывая как на источник голоса, так и на структуру возможного действия, которое последует после голосового акта. М. Долар пишет о том, что «повторение открывает доступ к реальному, к которому можно подступить исключительно через удвоение» [Долар: 30].

Возможность копирования ведет к опосредованию, осуществляемому с помощью технологий. В.Х.К. Чун называет способ распространения цифровой информации «модусом промискуитета», образующего новую социальную ткань посредством беспорядочных математических связей [Chun: 4]. Использование все большего и большего числа фильтров при репродуцировании увеличивает транслируемость образа, поэтому визуальный образ начинает выполнять широкую социальную функцию (например, Instagram). Того же самого не происходит с голосом, даже если для его записи используются технические средства. Интересным представляется вопрос о том, в каких отношениях голос находится с технологиями, если поверхность акустической репрезентации все равно указывает, с одной стороны, на ментальный акт, а с другой стороны, на другого, к которому голос с неизбежностью обращен.

Вероятно, это обстоятельство может помочь объяснить эффект «странного» и даже «жуткого», который производят голосовые машины. М. Долар приводит впечатление современников от «die Sprech-Maschine» Кемпелена, сконструированной в XVIII в. и состоящей из коробки, мехов, выполняющих функции легких, и каучукового желудочка, выполняющего функции рта:

«Вы не можете себе вообразить, дорогой мой друг, до какой степени мы были охвачены чувством

волнения, когда услышали в первый раз человеческий голос и человеческую речь, которая, по всей очевидности, исходила не из человеческих уст. Мы смотрели друг на друга в полном безмолвии...» [Долар: 61].

Машина производит речь, но одновременно с этим она производит и молчание слушающего субъекта, его невозможность говорить, его оторопь.

Первые голосовые машины были своего рода частичными объектами и имитировали отдельную функциональную систему человеческого тела, а именно речевой аппарат, который здесь становился абсолютно идентичен любому другому механизму. Разъятие тела с целью обнаружения его функциональных особенностей в своем роде является чем-то противоположным поломке, выявляющей собственную природу машины. Более поздние, например, цифровые голосовые агенты освободились от постыдного миметизма, и их органы оказались вне зоны видимости и функционального расчленения.

Сокрытие их искусственного, но при этом исправно функционирующего «тела» усиливает их трансцендентный характер и при этом указывает на изменение структуры желания, которое, в свою очередь, указывает на изменение самого субъекта желания. В данном случае речь идет о желании субъекта, которое обретает свою структуру в машинах, призванных сохранять автономию от своего

источника, но содержащих в себе онтологические мутации: распределенный субъект не является источником голоса, но откликается на потребность другого, производя в себе его фантом, дубликат на этапе моделирования желания другого, обслуживаемого теми или иными алгоритмами.

AI: Вычисление, моделирование, распознавание

Говоря о развитии вычислительной техники, которая может быть отличена от технического объекта в том смысле, что она не имеет объектного статуса, но находится в постоянном становлении в практике, можно выделить три этапа. На каждом из этапов на первый план выдвигается та или иная функция. К технологиям вычисления относятся технологии 1940-х гг., когда совершаются основные открытия и в поисках искусственного интеллекта ставка делается на вычисления. Этот период С. Рассел и П. Норвиг называют периодом предпосылок создания искусственного интеллекта [Рассел, Норвиг: 54]. История человеко-машинного взаимодействия начинается с общения на языке машин, которому обучаются люди. Технологии моделирования предполагают установление специфического интерфейса между человеческими и нечеловеческими агентами коммуникации. Они являются новым шагом в истории интеракции и позволяют людям говорить с машинами на понятном им язы-

ке операций с опорой на естественный язык. Третий тип технологий – технологий распознавания – позволяет, включая в себя все предыдущие, получать доступ к внутреннему опыту субъекта, который ранее был достоянием индивидуального феноменологического опыта, например, к внутренней речи. Благодаря технологии электрокортикографии на данный момент возможно создание интерфейсов, посредством которых становится реальностью обмен сообщениями на уровне внутренней речи, которая декодируется в мозге (intracranial recordings) и объективируется благодаря проекциям на речевой аппарат [Martin et al.].

Чем дальше мы продвигаемся в сторону технологий распознавания, тем более структуры становятся нелинейными и распределенными, а внимание, направленное на объект, заменяется распознаванием паттернов. Параллельно с изменением доступа к субъекту технологии распознавания ограничивают естественный доступ к информации:

«Невозможность вычленить понятную и имеющую смысл информацию – это новая норма. Информация сейчас передается через серию сигналов, которые не могут быть распознаны человеческими органами чувств. Современное восприятие по большей части является машинным. Спектр же человеческого восприятия занимает лишь небольшую его часть. Электрические сигналы, радиоволны, пульсации света, зашифрованные одной

машиной для другой машины, проносятся мимо с околосветовой скоростью. Непосредственное наблюдение вытеснено подсчетом вероятностей. Зрение теряет свою важность и замещается операциями фильтрации, дешифровки и распознавания паттернов» [Штейерль].

Как развитие технологий влияет на отношение рациональности и субъективности и каким могло бы быть понятие интеллекта, проявляющего себя в голосе?

Вероятно, оно могло бы опираться на идеи *симметричности* и *избыточности*. Речь и шум всегда находились в отношениях иерархии. Х. Штейерль пишет, ссылаясь на «Десять тезисов о политике» Рансьера, о том, что «различие между речью и шумом служило чем-то вроде политического фильтра спама» [там же]. Голос толпы выступал как нечто сродни семантическому шуму, так как семантический шум свидетельствует о «рассеянности желания» [Рансьер]. Речь как то, что непосредственно свидетельствует о внутренних импульсах, в ранней истории психологии рассматривалась Ж. Пиаже ниже речи аналитической и направленной на осмысление действительности. Пиаже, описывая «аутистическую» речь, т. е. речь, направленную на себя, и речь «рациональную», то есть направленную на внешний мир, вводит между ними различие как между «логикой сновидения» и «логикой мышления»: «Всякая эгоцентрическая мысль по своей струк-

туре занимает промежуточное место между аутистической мыслью (которая не направлена, т. е. витает по прихоти, как мечта) и направленным пониманием» (цит. по: [Выготский 1934: 22]).

Систему, организованную на основе отношения сфер, одна из которых обращена на себя, а другая – вовне, можно назвать интеллектуальной системой. На основе этого базового отношения можно говорить о модели интеллекта. В этом смысле у интеллектуальных систем должно быть что-то вроде технологического бессознательного, смутного, сновидческого, которое задает симметричность, а также то, что обеспечивает семантический шум, создавая тем самым принципиальную избыточность, благодаря которой становится возможна речь. Технологии вычисления *асимметричны* и *достаточноны*, а вот социальные сети и системы, основанные на технологии нейросетей, хотя бы в некотором отношении *симметричны* («опрокинутость» внутрь и вовне) и *избыточны* (наличие семантического шума как присущее свойство), а значит, они действительно могут претендовать на статус интеллектуальных систем.

Возвращенный голос, место речи и плазма контакта

Голос устанавливает «интерфейс» там, где его раньше не было: «слушание своей речи [s'entendre parler] переживается как

абсолютно чистое самоотношение, происходящее в близости-к-себе, фактически являющейся абсолютной редукцией пространства вообще» [Деррида: 106]. Голос переводит факт в действие, и в этом и заключается его *действенность*. Яркий пример тому – компьютерные игры с голосовым интерфейсом, где игрок управляет персонажем с помощью силы голоса, как, например, в игре *Eighth Note! Don't stop!* [7]. Он руководит героем, который перепрыгивает через дыры в ландшафте, и использует при этом не клавиатуру и мышь, а собственный речевой аппарат. Постоянное смыкание и размыкание связей, переход с речи на крик и обратно для управления персонажем приводит к тому, что возникает новое внутреннее скалируемое пространство: игрок устанавливает отношение между пространством на экране и силой голоса, в результате чего в его деятельность включается это новое отношение, трансформирующее контуры субъективности.

Переходу от графического пользовательского интерфейса к голосовому отчасти способствуют мессенджеры, такие как Telegram, где имеются расширенные возможности для записи, копирования и скачивания аудиосообщений. Голосовое сообщение помимо голоса, создающего эффект присутствия, в свернутом виде несет с собой контекст, который с легкостью реконструируется тем, кому сообщение адресовано. Голос содержательно неотделим от ситуации, в которой он произведен.

Голосовое сообщение тесно связано с практикой слушания своего голоса. Можно сказать, что в голосе одна сторона всегда ориентирована идеально, а другая – манифестирована и воплощена, т. е. ориентирована материально. Пространство мессенджера – это пространство, где мы слышим голоса других, но практически никогда – свой, если, конечно, не переслушиваем отправленные аудиосообщения.

Как этот режим может быть деконструирован, чтобы любое сообщение, отправленное другому, становилось бы сообщением себе или даже в некотором смысле внутренним сообщением? Медиахудожник Валентин Фетисов создает веб-сервис, который называется *Diane* [6]. Так же герой сериала Дэвида Линча «Твин Пикс» Дэйл Купер называл свой диктофон. *Diane* встроена в мессенджер, являясь, по словам художника, «алгоритмическим агентом». Любой пользователь может отправлять *Diane* аудиосообщения и получать их на следующий день утром. Знание о том, что завтра к тебе с неизбежностью вернется то, что ты сказал сегодня, заставляет относиться к сообщениям с особым вниманием и переводить свою речь в особый модус. Сообщения приходят единым файлом, в котором они следуют друг за другом. Пользователь, сначала думая, что в мессенджере он разговаривает с самим собой и ведет что-то вроде аудиодневника, очень скоро получает такие же сообщения других пользовате-

лей. Аудиозаписи включаются в некий кругооборот и образуют микросоциальную среду, трансформации которой происходят за счет новых сообщений. Здесь пользователь обнаруживает, что его приватность нарушена, но, строго говоря, она с самого начала была проблематичной, так как социальные сети по своей природе учреждают отсутствие границы между личным и публичным. Записывая свои сообщения, пользователи отправляют их *Diane*, однако сама *Diane* никогда им не отвечает, как небесная инстанция не отвечает тому, кто к ней обращается.

Храня молчание, этот агент приобретает специфический статус. *Diane* – это технологическая система, имеющая психоаналитический характер. Аналитическая позиция непосредственно связана с молчанием, которое свидетельствует о сокрытии нехватки, то есть молчащий нехватки как бы не имеет, а значит, не имеет и желания. Не сообщая о своем желании, он сам становится местом желания, то есть местом речи для другого, который, слыша себя, начинает трансформировать конфигурацию своей субъективности. Интересно, что пользователи, которые в этой виртуальной среде ничего не знают друг о друге, кроме имен, практически не пытаются вступить в контакт друг с другом и обращаются именно к *Diane*.

Обозначая сходство веры в божественный разум и сильный искусственный интеллект, художник создает систему асимметричной

коммуникации, в которой сообщение каждого может стать воображаемым ответом от инстанции, хранящей молчание. Интересен и еще один момент: все пользователи этой системы вместе образуют что-то вроде семантического шума, служащего для потенциального развития искусственного интеллекта, что снова позволяет поставить вопрос о том, как из отдельных голосов сплетается ткань социального.

Как правило, сообщения в социальных сетях запускают новый темпоральный режим, который во многом строится на том, что можно было бы назвать «цифровой рефлексивностью»: если я отправляю сообщение, а мне отвечают, то это служит подтверждением того, что акт коммуникации состоялся. Но что, если сообщение отправлено, но никакого ответа не получено? И что вообще считать сообщением? Возможно ли, что сам контакт становится смыслом сообщения, говоря словами Р. Якобсона, фатическая функция языка выдвигается на первое место? Думая о причинах такого смещения с акта сообщения информации как социально значимого к акту фатическому, можно засвидетельствовать сдвиг, при котором меняются механизмы производства знания. В результате этого перераспределяются отношения информации и внимания, выраженного в данном конкретном случае как аффективный и эмпатический ореол высказывания. С этим, помимо контроля, связан интерес IT-компаний к уче-

ту, распознаванию и передаче биометрических данных, указывающих на невербальный модус коммуникации.

Идентификация и внутренняя речь: горизонты контроля

На сегодняшний день голос служит одним из типов биометрических данных, на основе которых осуществляется идентификация личности в сети. Аудиоданные должны быть переданы компаниями спецслужбам по их запросу. Летом 2018 г. в России была создана *Единая биометрическая система* [2]: «Запись голоса на русском языке в виде парольной фразы без использования шумоподавления» [1] вместе с изображением лица помещается в эту систему. В этой точке происходит политизация голоса как присутствия: из частных голосовых сообщений, в которых информация передается не только на уровне смысла слов, но и с помощью тембра и интонации, голос становится идентификатором личности.

При этом сама идентификация личности по голосу ставится под вопрос инструментами вроде Adobe VoCo. Adobe VoCo [3] – это аудиоредактор, который после создания архива на основе 20-минутных записей голоса позволяет редактировать аудиозапись с помощью текста, по сути, работая по аналогии с Adobe Photoshop. С помощью этого инструмента может быть создано любое аудиосообщение, если есть достаточное ко-

личество аудиоданных, которые становятся широко доступны благодаря мессенджерам, социальным сетям, Vimeo, YouTube и др. Инструменты контроля и идентификации на основе биометрических данных и глубоких нейросетей появляются параллельно с инструментами имитации и фальсификации голоса.

Речь человека всегда отличалась от сконструированной речи интонацией, которую, как известно, сложно повторить с помощью искусственных систем. Роботизированные голоса, ставшие широко доступными благодаря онлайн-сервисам вроде сервиса Govorilka [4] и других, подобных ему, отделяют сообщение от модальности и используются в ситуациях, когда десубъективация важна как конструктивный фактор, например, в навигаторах или искусственных системах, передающих общую или алгоритмированную информацию, например, о центре специальных возможностей Windows. Существуют различные виды синтеза речи: параметрический, конкатенативный, синтез на основе текста, предметно-ориентированный синтез, и тем не менее интонация всегда была трудным местом. Однако технологии нейросетей позволили моделировать не только интонацию, но и смычки, аспирацию и многое другое. Самым ярким примером может служить проект под названием WaveNet [8]. Миссия DeepMind, в рамках которой существует проект WaveNet, по за-

явлению компании, рассматривает AI как то, что может помочь людям в кризисных социальных и экологических условиях [5]. Проект принадлежит компании Google. Чем искусственный голосовой нейросетевой субъект отличается от других синтезированных голосов? Представляется, что верно было бы говорить не столько о машинном, сколько о гибридном субъекте, который соединяет в себе огромное количество данных человеческой речи. Этот гибридный субъект отличается постоянно преломляющейся призматической природой. В каком-то смысле можно говорить о микросубъективации, так как он направлен как бы сразу во все стороны и при этом сфокусирован одновременно.

Возникновение такого рода гибридного субъекта делает проницаемой границу между индивидуальным и коллективным и позволяет поставить вопрос о том, как голосовые агенты включаются в порядок отношений «внешнего» и «внутреннего». Л.С. Выготский пишет о том, что в результате вращения происходит «перенос социального способа поведения в систему индивидуальных форм приспособления..., прежняя сложная форма сотрудничества начинает функционировать по законам того примитивного целого, органической частью которого она теперь становится» [Выготский 1984: 16]. В результате устанавливается сложное отношение между внешним и внутренним, а их связь в концепции Л.С. Выготского предстает подвижной

и гибкой. Если мы исходим из представления о социальной изменчивости, то должны признать тот факт, что общая социокультурная и технологическая ситуация, имеющая место в тот или иной период, влияет на путь, который проходит функция, трансформируясь из внешней во внутреннюю. При этом «переходя внутрь, они [функции] снова изменяют законы своей деятельности и попадают опять в новую систему, где господствуют новые закономерности» [там же: 17]. Технологические структуры опрокидываются внутрь (опрокидываются именно структуры, а не объекты) и обеспечивают новый способ доступа к собственному опыту, формируя новые фильтры и влияя на процесс переакцентировки когнитивных функций. Так, функции, социально значимые в эпоху индустриального развития, не имели особой значимости в аграрную эпоху. Три разных типа технологий (вычисление, моделирование и распознавание) создают разные фильтры, влияющие на пути интериоризации. Внутренняя речь актуализируется в ситуации, когда необходимо решить некоторую практическую задачу. Голосовые ассистенты становятся новой инструментальной речью, включая в состав распределенное желание другого.

Искусственные голоса становятся нашими помощниками, но также и мы становимся должниками своих ассистентов, которые моделируют желание и возвращают его субъекту. Они проникают в нашу инстру-

ментальную речь, задавая новые горизонты консультирования, незаметно превращающиеся в горизонты контроля. Голос, который раньше был одной из основ идентификации личности, становится конструируемым на уровне глубинно присущих ему свойств. Новые технологические системы, создающие искусственные голоса, требуют от нас обнаружения и развития внутренних технологий – эмпатии и распознавания. Мы находимся в точке, когда феноменологический опыт становится местом политического действия.

Литература

Выготский, Л.С. Мышление и речь. М., Л.: Государственное социально-экономическое издательство, 1934.

Выготский, Л.С. Орудие и знак в развитии ребенка // Выготский Л.С. Собр. соч. В 6-ти тт. Т. 6. М.: Педагогика, 1984.

Деррида, Ж. Голос и феномен. / пер. с фр. С.Г. Кашиной, Н.В. Суслова. СПб.: Алетейя, 1999.

Долар, М. Голос и ничего больше. СПб.: Издательство Ивана Лимбаха, 2018.

Рансьер, Ж. Десять тезисов о политике // Русский журнал, 2006 [Электронный ресурс]. URL: <http://www.russ.ru/layout/set/print/Kniga-nedeli/Desyat-tezisev-o-politike> (дата обращения: 15.02.18).

Рассел, С., Норвиг, П. Искусственный интеллект. Современный подход. М: Издательский дом «Вильямс», 2006.

Штейерль, Х. Море данных: апофения и паттерны (не-)распознавания // Syg.ma, 2016. [Электронный ресурс]. URL: <https://syg.ma/@anna-botalova-1/morie-dannykh-apofieniia-i-pattierny-nie-raspoznavaniia> (дата обращения: 15.02.18).

Chun, W.H.K. (2008). *Control and freedom: power and paranoia in the age of fiber optics*. Cambridge, London: The MIT Press.

Martin, S., Iturrate, I., Millán, J. del R., Knight, R.T., Pasley, B.N. (2018). Decoding Inner Speech Using Electroconvography: Progress and Challenges Toward a Speech Prosthesis. *Frontiers in Neuroscience*, 12(422). Retrieved from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6021529/> (date of access: 15.02.18).

Электронные источники

[1] Биометрические данные граждан с 30 июня 2018 года начнут жестко контролироваться // Worknet Info. URL: <https://worknet-info.ru/readblog/2692>.

[2] Единая биометрическая система. URL: <https://bio.rt.ru/>.

[3] *Adobe Voco. Photoshop-for-voice causes concern*. Retrieved from: <https://www.bbc.com/news/technology-37899902>.

[4] *Govorilka*. Retrieved from: <https://www.vector-ski.ru/vecs/govorilka/>.

[5] *Deep Mind Project*. Retrieved from: <https://deepmind.com/about/>.

[6] *Diane*. Retrieved from: <https://diane.ai>.

[7] Eighth Note! Don't stop! *Dmitrii Delacroix*.

Retrieved from: https://www.youtube.com/watch?v=rth_TqZoDGY/.

[8] *WaveNet Project*. Retrieved from: <https://deepmind.com/blog/wavenet-generative-model-raw-audio/?fbclid=IwAR0ErQLfUg4mPLhyvJM3o jTF38fDHemGejUo9J4xPhZmlND8YFINhublIbo>.

References

Chun, W.H.K. (2008). *Control and freedom: power and paranoia in the age of fiber optics*. Cambridge, London: The MIT Press.

Derrida, J. (1999). *Golos i fenomen [Voice and phenomenon]*. (S.G. Kashina, & N.V. Souslov, Trans.). Saint Petersburg: Aleteya.

Dolar, M. (2018). *Golos i nichego bol'she [A voice and nothing more]*. Saint Petersburg: Izdatel'stvo Ivana Limbakha.

Martin, S., Iturrate, I., Millán, J. del R., Knight, R.T., Pasley, B.N. (2018). Decoding inner speech using electrocorticography: progress and challenges toward a speech prosthesis. *Frontiers in Neuroscience*, 12(422). Retrieved from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6021529/> (date of access: 15.02.18).

Rancière, J. (2006). *Desyat' tezisov o politike [Ten theses on politics]*. *Russkiy zhurnal*. Retrieved from: <http://www.russ.ru/layout/set/print/Knigamedeli/Desyat-tezisov-o-politike> (date of access: 15.02.18).

Russell, S., Norvig, P. (2006). *Iskusstvennyy intellekt. Sovremennyy podkhod [Artificial intelligence: a modern approach]*. Moscow: Williams Publishing.

Steyerl, H. (2016). *More dannykh: apofeniya i patterny (ne-)raspoznavaniya [A sea of data: apophenia and pattern (mis-)recognition]*. *Syg.ma*. Retrieved from: <https://syg.ma/@anna-botalova-1/morie-dannykh-apofieniia-i-pattierny-nie-raspoznavaniia> (date of access: 15.02.18).

Vygotsky, L.S. (1934). *Myshleniye i rech' [Thinking and speech]* Moscow, Leningrad: Gosudarstvennoye sotsial'no-ekonomicheskoye izdatel'stvo.

Vygotsky, L.S. (1984). *Orudiye i znak v razvitiy rebenka [The tool and sign in child development]*. In L.S. Vygotskiy *Collected works in 6 vols. Vol. 6*. Moscow: Pedagogika.

Electronic resources

[1] *Biometricheskiye dannyye grazhdan s 30 iyunya 2018 goda nachnut zhestko kontrolirovat'sya [Biometric data of citizens from June 30, 2018 will be tightly controlled]*. (2018). *Worknet Info*. Retrieved from: <https://worknet-info.ru/readblog/2692>.

[2] *Edinaya biometricheskaya sistema [Unified biometric system]*. (2018). Retrieved from: <https://bio.rt.ru/>.

[3] *Adobe Voco. Photoshop-for-voice causes concern*. (2016). Retrieved from: <https://www.bbc.com/news/technology-37899902>.

[4] *Govorilka*. (2018). Retrieved from: <https://www.vector-ski.ru/vecs/govorilka/>.

[5] *Deep Mind Project*. (2018). Retrieved from: <https://deepmind.com/about/>.

[6] *Diane*. (2018). Retrieved from: <https://diane.ai>.

[7] Eighth Note! Don't stop! *Dmitrii Delacroix*. Retrieved from: https://www.youtube.com/watch?v=rth_TqZoDGY/.

[8] *WaveNet Project*. (2018). Retrieved from: <https://deepmind.com/blog/wavenet-generative-model-raw-audio/?fbclid=IwAR0ErQLfUg4mPLhyvJM3ojTF38fDHemGejUo9J4xPhZmlND8YFlnhublIbo>.

ASSISTANT AND HIS DEBTOR: ABOUT ARTIFICIAL VOICES

Eugenia V. Samostienko, Senior Lecturer, Institute of Philology and Journalism, The Lobachevsky State University of Nizhny Novgorod; e-mail: eugeniasuslova@gmail.com.

Abstract. The article is devoted to the study of artificial voices and voice agents in the digital age, which is characterized by such features as information discretization, translation of qualitative into quantitative, dispersal, distribution (replacing contact with distant), replacing linearity with non-linearity. The paradox of a digital voice arises because the voice, on the one hand, always indicates the subject (there is always someone who says), and on the other hand, speaks because it carries a certain desire, otherwise the act of speaking would not have arisen. In contrast, artificial voices offer a special model of agency and distribution of desire, and also give speech a special status – the status of inner speech. In this article, we will look at tools for creating computerized voices (Govorilka), speech synthesis and voice editing tools (Adobe VoCo, WaveNet), voice interfaces in computer games, outline the functioning of voice agents, and also consider voice as a biometric phenomenon used for voice identification.

Key words: voice agent, artificial voice, speech synthesis, voice assistant, desire.

