

На правах рукописи

Санникова Елена Георгиевна

**Восприятие и понимание
естественной и синтезированной речи**

Специальность 19.00.01 – Общая психология,
психология личности, история психологии

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата психологических наук

Екатеринбург – 2008

Диссертация выполнена на кафедре общей психологии и психологии личности Уральского государственного университета им. А.М. Горького

Научный руководитель	доктор психологических наук, Воронин Владимир Митрофанович
Официальные оппоненты	доктор психологических наук, профессор Прядеин Валерий Павлович кандидат психологических наук, Иванова Екатерина Сергеевна
Ведущая организация	ГОУ ВПО «Курганский государственный университет»

Защита состоится 25 июня 2008 г. в 15.00 часов на заседании диссертационного совета ДМ 212.286.07 при Уральском государственном университете им. А.М.Горького по адресу: 620083, г.Екатеринбург, пр. Ленина, 51, к. 248.

С диссертацией можно ознакомиться в библиотеке Уральского государственного университета им. А.М.Горького.

Автореферат разослан

мая 2008 г.

Ученый секретарь
диссертационного совета,
кандидат педагогических наук, доцент

Л.Г.Попова

общая характеристика работы

Актуальность исследования.

Способность говорить и воспринимать речь - важнейший аспект человеческого поведения. Аудиальная система человека играет решающую роль в такой, казалось бы, естественной и привычной, но по своей сути исключительно сложной форме человеческого общения, как устная коммуникация посредством языка. Благодаря когнитивной связи между голосовым аппаратом и слухом человек может генерировать и воспринимать речь, передавать другим сложную информацию, объем которой практически не ограничен.

Исследованию восприятия и понимания речи в её естественном варианте посвящено большое число работ как отечественных, так и зарубежных ученых (Л.А. Чистович, В.А. Кожевников, 1965, 1969; А.А. Леонтьев, 2003; И.А. Зимняя, 1989, 2001; Т.Н. Ушакова, 1999, 2003, 2004), G. Fant, 1964; D.Flanagan, 1968) и др. За рубежом интенсивно развиваются модели порождения речи, основанные на лингвистических концепциях, идущих от Н. Хомского (1957). Большое развитие получают психологические концепции восприятия и порождения речи, выполненные в рамках когнитивной психологии, и опирающиеся на идеи, представленные в трудах Дж. Миллера, Ю. Галантера, К. Прибрама (1965). В отечественной психологии большинство моделей речевосприятия и особенно речепорождения опираются на положение Л.С. Выготского (1956) об экстернизации мысли через внутреннюю речь к слову (А.Р. Лурия, 1975); А.А. Леонтьев, 2003). Получила также своё развитие концепция, основанная на деятельностном подходе (И.А. Зимняя, 2001), которая рассматривает речевую деятельность не только с точки зрения мотива, но и с точки зрения предмета, средств и способов деятельности. Наконец, в последние два десятилетия интенсивно развиваются в рамках коммуникативного психологического подхода «коммуникативные» модели речепорождения (Н.Д. Павлова, 2000; Т.Н. Ушакова, 1999).

В настоящее время в связи с развитием информационных технологий появляется возможность речевого общения человека с техническими системами, основанном как на анализе, так и на синтезе речевых сигналов. В этой связи важно знать, как на слух воспринимается синтезированная речь и насколько она отличается от естественной. Анализ степени научной разработанности темы диссертационного исследования позволяет сделать вывод о том, что только незначительное число работ зарубежных исследователей посвящено интересующей нас проблематике (D. Pisoni, 1983; Н. Nusbaum, 1983; В. Greene, 1984). Психологических же исследований с русскоязычными синтезаторами, позволяющих получить фундаментальную информацию о том, как человек воспринимает и понимает синтезированную речь, генерируемую по алгоритмам, каковы отличия синтезированной речи от естественной, в литературе не найдено.

Таким образом, обращение с нашей стороны к проблеме восприятия и понимания естественной и синтезированной речи обусловлено, во-первых, тем, что в современной жизни человеку с неизбежностью и всё возрастающей потребностью приходится вступать в речевое общение с современными техническими системами. Во-вторых, на основе сравнительной информации о восприятии и понимании двух видов речи, существенно расширяются наши знания о механизмах речепорождения и речевосприятия. Наконец, в-третьих, полученные знания послужат

совершенствованию самих систем речевого синтеза и их применению в разнообразных сферах человеческой деятельности.

Объект исследования: речевая деятельность человека.

Предмет исследования: сравнительный анализ восприятия и понимания естественной и синтезированной речи.

Цель исследования: изучение общих закономерностей и различий в восприятии и понимании естественной и синтезированной русскоязычной речи.

Задачи исследования:

1. Провести теоретический анализ и обобщить данные по разрабатываемой проблеме.

2. Определить факторы, влияющие на правильность распознавания синтезированной речи слушателем.

3. Сформировать комплекс наиболее информативных методик для сравнительного изучения восприятия естественной и синтезированной речи.

4. Экспериментально исследовать особенности восприятия русскоязычной синтезированной речи в сравнении с естественной и выявить отличия в разборчивости фонем и распознавании слов в предложениях, при принятии лексических решений, в латентности распознавания слов и словоподобных звуков.

5. Экспериментально обосновать справедливость предположения о том, что акустико-фонетическую структуру синтезированной речи труднее декодировать, чем естественную по тем же самым причинам, по каким трудно воспринимать естественную речь, предъявляемую на фоне шума.

6. Экспериментально исследовать возможность субъективного оценивания восприятия русскоязычной синтезированной речи.

7. Разработать и апробировать методику для количественной оценки понимания связного текста, произносимого с помощью синтезированной речи.

8. Экспериментально исследовать влияние утомления на понимание связного текста, произносимого с помощью синтезированной речи.

Гипотезами исследования являются следующие предположения:

1. На правильность распознавания синтезированной речи слушателем влияют ограничения, присущие системе обработки информации человеком, и акустико-фонетическая структура речевого сигнала.

2. Имеются определенные отличия в восприятии и понимании естественной и синтезированной речи как на фонематическом, так и на лексическом уровнях.

3. Синтезированную речь следует рассматривать как аналог «зашумленной» естественной при объяснении большей трудности ее перцептивной расшифровки.

4. В состоянии утомления за счет снижения общего запаса когнитивных ресурсов человека должны проявиться достоверные различия между пониманием повествовательных текстов, произносимых посредством синтезированной и естественной речи.

Теоретико-методологической основой диссертации являются теория деятельности (Л.С. Выготский, 1956, 1982; А.Н. Леонтьев, 1975), концепции современной когнитивной психологии, отечественной и зарубежной психолингвистики, лингвопсихологии и когнитивной науки в области восприятия речи (Н.И. Жинкин, 1956; И.А. Зимняя, 2001; Л.А. Чистович, 1965, 1969; А. Liberman, 1967; G.Fant, 1964; D.Flanagan, 1968; D.Pisoni, 1983) и в области понимания текстов (Н.И. Жинкин, 1956, 1982; И. А. Зимняя, 2001; А.А. Леонтьев,

2003; Т.Н. Ушакова, 2003; Kintsch W. & van Dijk T.A. 1978, 1988, 1998; Just M. A. & Carpenter P.A., 1980, 1987; Н. Хомский 1957, 1972, 1995; Thorndyke P.W., 1977).

Научная новизна исследования:

Первые в психологическом исследовании введены в научный оборот и изучены особенности восприятия русскоязычной речи, синтезированной по правилам. Впервые определены факторы, влияющие на правильность ее распознавания, а именно: лингвистическая структура сообщения; структура и качество речевого сигнала; ограничения, присущие когнитивной системе обработке информации, которой наделен человек; опыт и тренировка слушателя; конкретные требования и условия, связанные со спецификой решаемой задачи. Предложен новый подход к оценке понимания связанных текстов, основанный на пропозициональном анализе и позволяющий дать количественную оценку понимания.

Теоретическая значимость исследования заключается во введении в психологию речи изучения восприятия и понимания синтезированной по правилам речи. На этой основе могут моделироваться механизмы произношения, восприятия и понимания речи человеком.

Практическая значимость исследования заключается в том, что полученные результаты по аудиторской оценке качества речевых синтезаторов должны учитываться при разработке и внедрении новых систем синтеза речи. Это относится как к системам, применяющим синтезированную речь в быту и промышленности, так и к системам в образовании и медицине. Особо следует отметить практическую значимость полученных результатов для обучения инвалидов по зрению и слуху.

Надежность и достоверность результатов обеспечивается соблюдением общеметодологических принципов, адекватностью используемых методик, сочетанием количественного и качественного анализа, теоретических и эмпирических методов исследования.

Апробация работы. Результаты работы докладывались и обсуждались на кафедре общей психологии и психологии личности Уральского государственного университета, Екатеринбургского городского центра «Здоровье». Данные исследования включены в учебные курсы: «Актуальные проблемы общей психологии», «Компьютерные технологии в психологии и образовании» для магистров - психологов первого года обучения.

По теме диссертации опубликовано пять работ общим объемом 4 п.л., в том числе одна статья в ведущем рецензируемом научном журнале, рекомендованном ВАК.

Положения, выносимые на защиту.

1. На правильность распознавания человеком русскоязычной синтезированной по правилам речи влияет следующий ряд факторов: 1) конкретные требования и условия, связанные со спецификой решаемой задачи; 2) ограничения, присущие системе обработки информации, которой наделен человек; 3) опыт и тренировка слушателя; 4) лингвистическая структура сообщения и 5) структура и качество речевого сигнала.

2. Восприятие синтезированной речи по сравнению с естественной имеет определенные различия, вызванные большей трудностью расшифровки синтезированной речи.

3. Различия проявляются: в разборчивости фонем и распознавании слов в предложениях, при принятии лексических решений, в латентности распознавания слов и словоподобных звуков.

4. Акустико-фонетическую структуру синтезированной речи труднее декодировать, чем естественную, по тем же самым причинам, по которым трудно воспринимать естественную речь, предъявляемую на фоне шума,- наличие маскирующего шума приводит к искажениям акустических признаков фонем или ухудшению акустического качества фонем. В соответствии с этой точкой зрения синтезированная речь по параметрам близка к естественной речи, однако по сравнению с ней является менее четко выраженной.

5. На основе пропозициональной структурной схемы рассказа П. Торндайка возможно построение метода количественной оценки понимания повествовательных текстов, произносимых посредством синтезированной речи.

Структура и объем работы.

Диссертация состоит из введения, трех глав, выводов, заключения (общий объем составляет 155 страниц машинописного текста); списка литературы, включающего 188 наименований, в том числе 69 на иностранных языках и приложений. Работа содержит 8 рисунков, 10 таблиц.

Основное содержание работы.

Во введении обосновывается актуальность темы диссертации, сформулированы цель и задачи работы, объект и предмет исследования, научная новизна, теоретическая и практическая значимость, представлены гипотезы и положения, выносимые на защиту.

В главе 1 «Теоретический анализ проблемы восприятия и понимания речи» представлен анализ работ отечественных и зарубежных ученых, посвященных изучению данной проблемы.

Первый параграф главы посвящен обзору работ по восприятию естественной речи, основанному на взаимодействии большого числа сложных психологических факторов. При анализе особый интерес представляли те эксперименты, процедуры и условия восприятия речи, в которых различные ее аспекты либо вовсе отсутствовали, либо были весьма существенно искажены или смазаны специально для того, чтобы можно было определить их влияние на разборчивость речи. К таковым относятся: «срезание частот» в классической работе Френча и Стейнберга (French N.R., & Steinberg, 1947), показавших, что в определенных пределах стимульная информация, критически важная для восприятия речи, не «привязана» ни к одному частотному диапазону; значение речевого контекста, которое исследовалось в экспериментах Bashford J. A., Riener K.R., & Warren (1992), получивших название «эффект восстановления фонемы»; влияние контекста на восприятие отдельных слов (Pollack, 1964); влияние контекста на восприятие фрагментов слова и гласных звуков, которое изучалось Rakerd B. (1984), показавшим, что изолированные гласные звуки воспринимаются менее лингвистично, чем когда они находятся в окружении согласных (консонантный эффект).

В контексте нашего исследования существенный интерес представляют также данные по выявлению субъективной вероятностной оценки воспринимаемого речевого сигнала в зависимости от того уровня языковой иерархии, к которому он принадлежит. Как известно, зависимость восприятия каждой конкретной речевой единицы от субъективной оценки вероятности ее появления устанавливается не

только на уровне фразы, но и на всех остальных уровнях языковой иерархии (И.А. Зимняя, 1971). Исключением не является и восприятие изолированных звуков, потому что и на этом уровне уже есть вероятностная оценка отнесения данного звука в зону того или иного эталона — звукотипа.

По нашему мнению, вероятностный характер восприятия будет проявляться не только по мере усложнения структуры речевого сигнала и тех связей, которые объединяют его части в целое — звуки в слог, слоги в слово и слова во фразы, а фразы в текст, как считает И.А. Зимняя (2001), но и при получении слушателем упрощенных речевых сигналов, звучание которых определяется лишь ограниченным подмножеством всего множества акустических параметров, используемых для передачи фонетической информации. Исходя из общепсихологической закономерности восприятия соотношения частей и целого, можно предположить, что чем шире этот контекст, чем сложнее структура целого, тем менее зависит восприятие от точности физической характеристики отдельных элементов или частей целого и тем больше оно будет обуславливаться вероятностной оценкой всего сигнала. Принимая какой-либо звуковой сигнал, человек сразу же выдвигает гипотезу о принадлежности этой части к какому-то целому, вероятность реализации которой основывается на широком контексте и в значительной мере на частотности этого целого в алфавите данного языка.

Из нашего обзора следует, что нельзя прийти к одному приемлемому и достаточно полному объяснению восприятия речи. С большой долей вероятности можно констатировать, что восприятие речи, как и другие сложные формы деятельности и способности человека, базируется не на одном, а на нескольких механизмах. В любом случае не должно быть сомнений в том, что восприятие речи далеко выходит за рамки простого анализа физических свойств речевых звуков.

Во втором параграфе главы представлен анализ работ по пониманию речи. Этой проблеме в отечественной литературе посвящено большое количество исследований в психологии, психолингвистике, когнитивной науке, философии и других гуманитарных дисциплинах (А.А. Брудный, 1975; Л.С. Выготский, 1956, 1996; В.Н. Дружинин, 2001; Н.И. Жинкин, 1956; А.А. Залевская, 2005; И.А. Зимняя, 2001; В.В. Знаков, 1986; А.А. Леонтьев, 2003; А.Р. Лурия, 1975; Н.А. Рубакин, 2006; А.А. Смирнов, 1966 и др.). Понимание, как справедливо отмечает Л.С. Цветкова (1995), является одним из самых сложных предметов психологического исследования ввиду недостаточной разработанности методов исследования, трудности выбора в эксперименте индикаторов и критериев понимания, а также сложности самого феномена «понимание», отсутствии его однозначного определения.

Ещё более малоизученной до настоящего времени предстаёт проблема оценивания результатов понимания (С.В. Шаповал, 2006). Это положение отмечается и в «Концепции образовательной области «Филология» (2000): «До сих пор не выработаны научно обоснованные критерии оценки знаний, умений и навыков учащихся...». Поэтому теоретический анализ отечественных и зарубежных работ по проблеме понимания проводится нами с целью определения теоретических подходов, на основе которых можно было бы построить метод количественного оценивания понимания повествовательных текстов, произносимых с помощью естественной и синтезированной речи.

Одним из основополагающих тезисов в отечественной психологии речи является утверждение, что восприятие и понимание образуют единый процесс (С.Л.

Рубинштейн, 1940; В.А. Артемов, 1954; Н.И. Жинкин, 1956; А.Н. Соколов, 1968; Б.В. Беляев, 1965 и др.). Для наших целей ближе точка зрения, которой придерживается И.А. Зимняя (1989), разграничивающая процессуальную сторону понимания, называемую осмыслением, и результативную его сторону, то есть внешний результат - ответное высказывание слушателя. В диссертации разбираются подходы В.А. Артемова (1954), С. Фессенден (1955), И.А. Зимней (1989, 2001). Значимой для образовательной практики является концепция З.И. Клычниковой (1973), которая не только выделяет четыре группы категорий смысловой информации и в соответствии с ними четыре плана текста и семь уровней его понимания в процессе чтения, но и показатели понимания текста, в качестве которых, по мнению автора, выступают способы проверки понимания. Однако подход не дает полного ответа на вопрос о том, каковы же оперативные критерия понимания речевого сообщения в процессе смыслового, особенно слухового восприятия. Ответ на этот нерешенный вопрос психологии речевого восприятия, по мнению И.А. Зимней (1989), следует искать в характере взаимоотношения уровней понимания и последующего воспроизведения воспринятого речевого сообщения. Продуктивный, на наш взгляд, подход к определению уровней понимания (с точки зрения меры осознания слушателем своей деятельности, то есть с позиции отчетливости понимания) был предложен А.А. Смирновым (1966), но и он не может быть положен в основу построения метода количественной оценки понимания.

Среди зарубежных теорий, прежде всего, нами выделяются психолингвистические, базирующиеся на фундаментальной гипотезе Хомского-Миллера (N. Homsky, G. Miller) о том, что понимание предполагает переход от поверхностной к глубинной структуре, или так называемой «ядерной репрезентации». Несмотря на то, что в последнее время произошли радикальные изменения во взглядах Хомского (N. Homsky, 1995) относительно представлений о глубинной и поверхностной структурах, его предположение о центральной роли синтаксиса в процессе понимания осталось неизменным.

Идеи психолингвистики и когнитивной науки послужили основой при построении моделей глобальной организации знаний и познавательных процессов, что, несомненно, явилось новым этапом в развитии когнитивного подхода (D. Anderson, G. Bower, 1985, 2002; P. Lindsay, D. Norman, D. Rumelhart, 1972, 1981; R. Schank, R. Abelson, 1977). Общим недостатком глобальных моделей является то обстоятельство, что при их построении структура репрезентации знаний выбирается без должного учета психологических данных, хотя успешность этой работы определяется не столько мощностью формального аппарата, сколько учетом особенностей организации повседневных форм активности человека.

Среди психологических моделей когнитивной психологии, которые сосредотачивают свое внимание именно на человеке и его процессах понимания текста, выделяются три модели: Just M.A. & Carpenter P.A., 1980, 1987; W. Kintsch & van Dijk, 1978, 1988, 1998; P. Thorndyke, 1977. Центральное место во всех трех моделях занимает понятие *пропозиции* как элементарной единицы знания. Модель Just & Carpenter основанная на эмпирическом измерении путем регистрации движений глаз, позволяет делать конкретные предсказания о деятельности чтения. В целом эта модель хорошо объясняет процесс понимания на широкой выборке испытуемых (87% дисперсии), хотя анализ индивидуальных данных показал, что корреляция между ними и предсказаниями модели меняется в диапазоне от 0,998 до -0,887, что свидетельствует о множестве индивидуально используемых средств

познавательной активности, не всегда сводящимся к дискретным логическим функциям. К числу наиболее явных недостатков этого подхода относятся: игнорирование контекста и схематической организации знания, которая несомненно играет важную роль в процессах понимания. Нужно также учесть, что эта модель, основанная на регистрации движения глаз, не работает, когда речь идет о слушании, а не о чтении.

В этой связи в работе подробно анализируется модель У. Кинча и ван Дийка (1988, 1998), которые сделали попытку учесть не только контекст, но и ситуацию переработки информации в целом. С целью улучшения понимания данной модели в работе приводятся теоретические обоснования ее и операционализируются понятия, которые необходимы для объяснения процесса понимания. Нами подчеркивается, что «текст» У. Кинча и ван Дийка понимается как весь поток информации, воспринимаемый человеком, т.е. в это понятие включено слушание и рассказывание, чтение и письмо, даже передача и восприятие невербальной информации во время коммуникации. Вместо традиционной структурной модели понимания У. Кинч и ван Дийк оперируют стратегической моделью, подразумевая под стратегиями эффективные рабочие гипотезы относительно правильной структуры и значения фрагмента текста. Стратегический анализ зависит не только от текстуальных характеристик, но и от характеристик пользователей языка, его целей и знаний о мире. Основным свойством модели У. Кинча и ван Дийка является предположение, что понимание текста подразумевает не только представление текстовой базы в эпизодической памяти (понятие, введенное в психологию Э. Тульвингом), но также активацию, обновление и другие способы функционирования так называемой ситуационной модели в эпизодической памяти.

Давая оценку практическому применению модели, нами отмечается, что модель У. Кинча и ван Дийка опирается на извлечение из текста пропозициональной информации, отсюда она позволяет исследователям делать точные предсказания о запоминаемости конкретных типов информации. Разработанная этими авторами модель в отличие от субъективного метода, используемого ранее Бартлетом (1958), – согласуется с научной методологией, принятой в современной психологии.

Идеи У. Кинча и ван Дийка послужили основой для нового подхода к текстам – представлению повествовательной информации в виде, когда более важные высказывания поддерживаются менее важными. Именно с такой позиции, отмечается в диссертации, подошел к изучению структуры повествования П. Торндайк. Он создал грамматику рассказа, сообразно определенным правилам и на примере рассказа «Остров круга» выделил 4 уровня расположения пропозиций, предполагая, что пропозиции более высокого уровня важнее для понимания рассказа, чем пропозиции более низкого уровня, что подтверждает структурный характер памяти на повествовательные тексты.

Из экспериментов П. Торндайка можно сделать некоторые выводы о том, как разум человека кодирует, обрабатывает, хранит и запоминает повествования (Солсо, 1996; Величковский, 2006). Во-первых, повествования строятся по определенным закономерностям и имеют свою структуру, что в свою очередь отражает склонность читателя кодировать и структурировать литературный материал. Во-вторых, повествования можно разделить на их структурные компоненты и выделить среди них элементы более высокого и более низкого уровня. Затем эти правила описания можно использовать для выделения основных элементов (обработки рассказа), учитывая, что элементы более высокого уровня в иерархии имеют большую

важность, чем элементы более низкого уровня. Это послужило основанием использования нами пропозициональной схемы П. Торндайка при разработке математической модели оценивания и метода количественной оценки понимания естественной и синтезированной речи,

В главе 2 «Синтез речи - новый этап в экспериментальном исследовании её восприятия и понимания» рассматриваются основные классы систем с речевым выходом - системы речевого кодирования, использующие фиксированное число параметров для воспроизведения сравнительно ограниченного набора высказываний, и системы синтеза речи, генерируемой по алгоритмам, обладающие способностью преобразовывать в речь любой текст. На основе общих концептуальных положений когнитивной психологии, а также на основании анализа, проведенного в первой главе, определяется ряд факторов, влияющих на правильность распознавания речи слушателем:

1) конкретные требования и условия, связанные со спецификой решаемой задачи; 2) ограничения, присущие системе обработки информации, которой наделен человек; 3) опыт и тренировка слушателя; 4) лингвистическая структура сообщения; 5) структура и качество речевого сигнала.

В работе подробно разбирается каждый из этих факторов, что создает методологическую основу для последующих экспериментов.

Ошибки в системе «текст - речь», которые влияют на общую разборчивость речи, могут происходить, в основном, в трех случаях: 1) при переводе букв и буквосочетаний в звуки; 2) при вычислении и воспроизведении супraseгментной структуры; 3) в правилах фонетического исполнения, когда внутреннее представление фонем и слогов преобразуется в звуковые колебания. Эти ошибки приводят к большей трудности расшифровки синтезированной речи по сравнению с естественной.

Можно предложить *две гипотезы, объясняющие особенности перцептивной расшифровки синтезированной речи*. Одна из них предполагает, что синтезированная речь просто аналогична «зашумленной» естественной речи (мы придерживаемся этой гипотезы). Иными словами акустико-фонетическую структуру синтезированной речи труднее декодировать, чем естественную по тем же самым причинам, по каким трудно воспринимать естественную речь, предъявляемую на фоне шума, - наличие маскирующего шума приводит к искажениям акустических признаков фонем или ухудшению акустического качества фонем. В соответствии с этой точкой зрения синтезированная речь по параметрам близка к естественной речи, однако по сравнению с естественной является менее четко выраженной. Согласно другой гипотезе, синтезированная речь вовсе не напоминает «зашумленную» или нечеткую естественную речь, а может рассматриваться как «перцептивно новая», обедненная по сравнению с естественной речью. С этой точки зрения, как считают D. Pisoni и его соавторы (1984), синтезированная речь принципиально отличается от естественной как в количественном, так и в качественном отношении, поскольку слабо представлены (или вообще отсутствуют) важные акустические признаки, по которым производится распознавание. Поэтому мы сочли необходимым экспериментально выяснить справедливость той или иной гипотезы.

В параграфе «Организация и методы исследования восприятия и понимания синтезированной речи» даётся характеристика выборки испытуемых, описываются отобранные и разработанные методики, а также процедуры статистического анализа данных. На различных этапах экспериментального исследования, разработки и

апробации методик изучения восприятия и понимания естественной и синтезированной речи приняли участие 220 человек в возрасте от 17 до 24 лет. В состав выборки вошли студенты и учащиеся старших классов средней школы. Главным критерием отбора испытуемых было нормальное состояние слуха, которое выяснялось посредством тональной и речевой аудиометрии.

При проведении экспериментов с синтезированной речью применялись четыре типа русскоязычных синтезаторов: с упрощенным синтезом, Speaking Mouse Home, TextAloud, Govorilka text-to-speech software и один англоязычный синтезатор Speech Sinthezier 5.0; компьютерное программное обеспечение для работы с речевыми сигналами WaveLab 3.03. Исползованные методики: для исследования разборчивости фонем применялся адаптированный нами американский модифицированный райм-тест (MPT) в двух процедурах: с ограниченным (шесть слов) и неограниченным набором вариантов; для выяснения большей трудности расшифровки синтезированной речи использовались метод стробирования сигнала и метод смешения слогов из согласных и гласных; для субъективного оценивания адаптирована анкета международной комиссии по синтетической телефонии (Женева, 1994); для количественной оценки понимания синтезированной речи разработана авторская методика оценки.

Стимульный материал для тестирования готовился следующим образом. Все слоги, слова, фразы естественной речи были произнесены диктором в студийных условиях. Запись производилась на цифровые носители и была сохранена в формате WAVE PCM с частотой дискретизации 22050 гц и разрядностью 16 бит. Фразы были произнесены со средним темпом речи, с небольшими вариациями частоты основного тона, с интонацией перечисления. Длительность пауз между словами была приведена к значению 200 мс. Диапазон частоты основного тона составил 75 – 125 гц.

В соответствии с задачами исследования были применены различные средства математической обработки данных. При экспериментальном выявлении особенностей восприятия русскоязычной синтезированной речи в условиях независимых выборок были применены критерий углового преобразования Фишера и параметрический критерий Стьюдента. Экспериментальная верификация гипотезы и выяснение возможностей субъективного оценивания восприятия синтезированной речи осуществлялись с привлечением однофакторного дисперсионного анализа. Для исследования количественной оценки понимания повествовательных текстов в условиях работоспособности и утомления были применены параметрический критерий Стьюдента и корреляционный анализ Пирсона.

Третья глава «Экспериментальные исследования восприятия и понимания естественной и синтезированной речи» посвящена изучению различий в восприятии и понимании естественной и синтезированной речи.

Экспериментальные исследования проводились в нескольких направлениях: в выявлении различий в восприятии синтезированной речи по сравнению с естественной на уровне фонем и слов в осмысленных и бессмысленных предложениях, а также при принятии лексических решений; в выяснении природы более трудной перцептивной расшифровки синтезированной речи по сравнению с естественной; в определении степени сходства синтезированной речи с естественной при субъективном оценивании; в экспериментальной проверке разработанного метода количественной оценки понимания повествовательных текстов.

В экспериментах *по разборчивости фонем* использовался метод, называемый модифицированным райм-тестом (МРТ). Сейчас этот тест принят в качестве основного для оценки сегментной разборчивости синтезированной речи за рубежом.

При МРТ слушатель должен идентифицировать одиночные слова, делая выбор при ответе из предлагаемых ему шести слов, которые отличаются друг от друга одной буквой в начале либо в конце слова. Примером являются односложные слова типа «согласная - гласная - согласная» (СГС): МИР - ТИР - ЛИР; РИС - РИМ; ДОМ - КОМ; МОР - СОР; СОР - СОК. МРТ позволяет оценить эффективность идентификации слушателем либо начальной, либо конечной фонемы в наборе произнесенных слов.

Помимо стандартной процедуры МРТ с ограниченным набором вариантов, из которых производится выбор, нами были проведены также испытания на основе МРТ при неограниченном выборе. Согласно этой процедуре в каждой экспериментальной попытке слушатель должен был записать то слово, которое он услышал. Данная методика позволяет оценить правильность распознавания в условиях, когда ограничения, налагаемые на возможные ответы, минимальны (ответом может быть любое известное слушателю слово типа «согласная -гласная - согласная», а не одно из шести предлагаемых слов, как в стандартной процедуре МРТ). Это позволяет получить также сведения относительно разборчивости гласных, что невозможно в методике МРТ с выбором из ограниченного числа возможных ответов. Мы провели исследования с применением МРТ с использованием 4-х русскоязычных синтезаторов.

Первый синтезатор осуществлял упрощенный синтез, второй - Speaking Mouse Home являлся двуязычным (русский, английский), десятиголосовым синтезатором. Третий синтезатор - Govorilka text-to-speech software (голос Digalo Russian Nicolai) и четвертый – TextAloud (голоса «Николай», «Катерина»). Общее число испытуемых составило 100 человек (две независимых выборки по 50 испытуемых в каждой).

В наших экспериментах при ограниченном наборе слов (шесть слов в наборе) естественная речь правильно была опознана в 99,8% случаев предъявления. Синтезированная речь при использовании упрощенного синтезатора была правильно опознана в 75% случаев, а при использовании синтезатора Speaking Mouse Home голосом «Диктор» в 95,5% случаев предъявления. При использовании Govorilka text-to-speech software опознание составило 94.2%, при использовании TextAloud («Николай») синтезированная речь была правильно опознана в 96.2% предъявлений, а при использовании TextAloud («Катерина») была правильно идентифицирована в 96.8% предъявлений.

При использовании МРТ с неограниченным набором слов показатели выглядят следующими образом:

естественная речь - 98,3%, упрощенный синтезатор – 67,0%, Speaking Mouse Home («Диктор») - 92,3%, Speaking Mouse Home («Mouse») - 89,2%, Speaking Mouse Home («King Kong») - 87,7%, Govorilka text-to-speech software – 91,3%, TextAloud («Николай») – 93,5%, TextAloud («Катерина») – 94,5%.

Эти результаты, полученные в одних и тех же лабораторных условиях при строгом соблюдении неизменности условий эксперимента, показали, что системы «текст - речь» существенно отличаются друг от друга. На наш взгляд, эти различия напрямую связаны с глубиной фундаментальных исследований, проводившихся с целью разработки правил фонетического исполнения для этих систем с речевым ответом.

Сравнение восприятия речи, синтезированной различными системами «текст - речь» в двух вариантах МРТ (с ограниченным и неограниченным выбором), и восприятия естественной речи позволяет оценить, насколько слушатель опирается на ограничения, налагаемые на набор ответов.

При выполнении теста МРТ с неограниченным выбором число правильно идентифицированных слов естественной речи в наших экспериментах составило 98.3%, тогда как при ограниченном выборе - 99,8% (у американских авторов - 97,2%, и 99,4% соответственно). Отсюда следует вывод, что правильность распознавания естественной речи даже при отсутствии ограничений на набор ответов была выше, чем речи, синтезированной с помощью любой системы «текст - речь» при ограниченном числе возможных ответов.

Полученные нами результаты указывают на наличие сильной взаимосвязи между разборчивостью речи при выполнении МРТ в двух вариантах - с ограниченным и неограниченным выбором. Несмотря на то, что порядок разборчивости для этих двух вариантов МРТ остается одинаковым, не вызывает сомнений, что по мере того, как речь становится менее разборчивой, слушатели, пытаясь распознать слова, все больше полагаются на ограничения, наложенные на возможные варианты ответа.

Для того, чтобы оценить влияние лингвистических закономерностей на правильность восприятия речи, мы провели сравнение правильности *распознавания слов в двух типах предложений*. Общее число испытуемых составило 100 человек (две независимых выборки по 50 испытуемых в каждой). Предложения первого типа были осмысленными и синтаксически корректными. Были взяты гарвардские психоакустические предложения и переведены на русский. Пример такого предложения: Add salt before you fry the egg. (Посолите яичницу заранее).

Предложения второго типа были корректными синтаксически, но неправильными семантически (синтаксические предложения Хаскинса), так же переведены на русский. Такие испытательные предложения синтаксически построены как обычные предложения, но не имеют смысла. Пример такого предложения: The old farm cost the blood. (Старая ферма стоит крови). Сравнив правильность распознавания слов в двух этих типах предложений, можно оценить влияние, которое оказывает на распознавание слов смысл предложения и лингвистические закономерности. Результаты наших экспериментальных исследований с тремя типами речевых синтезаторов и естественной речью отражены в таблице 1. Для оценки достоверности различий был применен критерий углового преобразования Фишера. Из таблицы видно, что достоверные различия получены для всех типов синтезаторов (стандартное значение критерия Фишера равно 1.64 для $p=0.05$), недостоверные - для естественной речи.

Таблица 1

Процент правильно опознанных слов в осмысленных и семантически бессмысленных предложениях в наших экспериментах

Речь	Тип предложения		Критерий Фишера
	Осмысленное в %	Бессмысленное в %	

Естественная	99,5	98,2	0.52 p<0.05
Упрощенный синтезатор	86,2	72,1	1.75 p<0.05
TextAloud «Катерина»	96,7	87,2	1.80 p<0.05
Speaking Mouse Home «Диктор»	95,5	85,0	1.84 p<0.05

Для сравнения в таблице 2 приведены данные, полученные в экспериментах американских авторов с англоязычными синтезаторами. Результаты проведения МРТ и экспериментов по идентификации слов естественной и синтезированной речи показывают, что синтезированная речь несколько менее разборчива, чем естественная.

Таблица
2

Процент правильно опознанных слов в осмысленных и семантически бессмысленных предложениях в экспериментах американских авторов

Речь	Тип предложения	
	Осмысленное в %	Бессмысленное в %
Естественная	99,2	97,7
MITalk-79	93,3	79,7
DEC Paul	95,3	86,8
DEC Betty	90,5	75,1

Кроме того, эксперименты показали, когда синтезированная речь становится все менее разборчивой, слушатели начинают все больше опираться на лингвистические правила и на ограничения круга возможных ответов, что помогает идентификации слов. Однако эксперименты не объясняют различий в восприятии естественной и синтезированной речи, их роль состояла просто в выявлении и описании этих принципиальных различий по идентификации слов.

Чтобы изучить различия в перцептивной обработке естественной и синтезированной речи, была выполнена серия экспериментов по измерению времени для **распознавания слушателем слов и словоподобных сочетаний** звуков, произносимых диктором и системой «текст - речь». В этих экспериментах выяснялось, сколько времени нужно слушателю, чтобы идентифицировать одиночное слово, и как на процесс распознавания слов влияет качество акустико-фонетической информации в сигнале. Для измерения продолжительности процесса

распознавания, мы воспользовались задачей принятия лексических решений. В экспериментах применялись русскоязычные синтезаторы Speaking Mouse Home («Диктор»), TextAloud («Катерина»). Слушателю в каждой экспериментальной попытке предъявлялось либо одиночное слово, либо стимульная словоподобная единица. Каждый слушатель должен был как можно быстрее и точнее классифицировать стимульную единицу как «слово» и «неслово», нажимая одну из двух кнопок на блоке ответов, соединенном с персональной ЭВМ. Общее число испытуемых составило 100 человек (две независимых выборки по 50 испытуемых в каждой). Оценка достоверности осуществлялась с помощью параметрического критерия Стьюдента, стандартное значение которого для нашей выборки при $p=0.01$ составляет 2.62.

Слушатели реагировали на произнесенные диктором слова русского языка (978 мс) и «неслова» (1008 мс) быстрее, чем на слова и «неслова», синтезированные машиной (1096 и 1203 мс соответственно). В среднем время реакции на синтезированную речь на 137 мс превышало время реакции на естественную речь ($t=3.54$, $p<0.01$). Полученные данные указывают на существование двух важных различий в восприятии естественной и синтезированной речи. Во-первых, восприятие синтезированной речи требует больше умственных усилий, чем восприятие естественной речи. Во-вторых, различия в латентности наблюдались в равной степени как для слов, так и для «неслов», и поэтому они не зависят от лексического статуса стимула. Как и в эксперименте с принятием лексических решений, слушатели заметно больше времени затрачивали на называние синтезированных тестовых стимулов, чем на называние естественных стимулов. Валидное различие опять-таки наблюдалось как для слов, так и для «неслов» ($t=3.2$, $p<0.01$). Отсюда следует, что дополнительные усилия при обработке информации, по-видимому, связаны с процессом выделения из сигнала акустико-фонетической информации, а не с процессом идентификации слов в лексиконе. Резюмируя, можно сказать, что совокупность полученных результатов позволяет предположить, что перцептивные процессы по расшифровке синтетической речи требуют больших умственных «усилий» или ресурсов, чем процессы расшифровки естественной речи.

Для выяснения природы более трудной перцептивной расшифровки синтезированной речи по сравнению с естественной были поставлены *эксперименты со стробированием сигнала и «защумленной»* естественной речью. Анализировались особенности восприятия слова в зависимости от степени стробирования сигнала и сравнивались временные параметры восприятия стробированных слов, предъявляемых в двух форматах: слова синтезированной речи, слова естественной речи. При проведении исследования стробировались слова, состоящие из пяти и трех слогов, при помощи программы WAVELAB 3.0. Сигнал стробировался от 20% до 50% длительности исходного слова. Для генерации слов синтезированной речи использовался синтезатор TextAloud. В качестве аудиторов выступили 40 человек (мужчины в возрасте от 17 до 22 лет) с нормальным состоянием слуха и опытом работы с синтезированной речью. Испытуемым предлагалось послушать стробированный сигнал и воспроизвести слово целиком, записав его на бланке. При этом работа проводилась в два этапа. На первом этапе испытуемым предлагалось прослушать стробированные слова, состоящие из 3 и 5 слогов, предъявляемые в формате синтезированной речи. На следующем этапе испытуемым предлагалось прослушать стробированные слова, состоящие из 3 и 5 слогов, предъявляемые в формате естественной речи. Для проверки достоверности

различий нами использовался однофакторный дисперсионный анализ и F-критерий. При обработке данных были получены следующие значения F-критерия; при сравнении стробированных слов, состоящих из 3 слогов, предъявляемых в формате синтезированной и естественной речи $F=0.06$; при сравнении стробированных слов, состоящих из 5 слогов, предъявляемых в формате синтезированной и естественной речи $F=1.36$; при сравнении стробированных слов, состоящих из 3 и 5 слогов, предъявляемых в формате синтезированной речи $F=47.73$; при сравнении стробированных слов, состоящих из 3 и 5 слогов, предъявляемых в формате естественной речи $F=47.21$. При этом стандартное значение F-критерия для наших условий равно 4.41. Полученные результаты свидетельствуют о существенных отличиях в восприятии стробированных слов, состоящих из 3 и 5 слогов одного формата предъявления. Эти результаты, полученные в ходе эксперимента, являются достоверными. Вместе с тем, анализ полученных результатов по оценке понимания стробированных слов, состоящих из 3 и 5 слогов, свидетельствует о несущественных отличиях синтезированной речи от естественной. Отличия, полученные в ходе эксперимента, не являются достоверными. В данном случае факт отсутствия достоверных отличий можно объяснить тем, что перед проведением эксперимента испытуемые проходили тренировку, прослушивая с помощью компьютерной программы TextAloud учебные тексты. Можно предположить, что, приобретя определенный опыт, испытуемые смогли преодолеть некоторые из ограничений, которые накладывает на восприятие синтезированной речь.

При **исследовании смещения слогов** в качестве аудиторов выступили 40 человек (мужчины в возрасте от 17 до 22 лет) с нормальным состоянием слуха и опытом работы с синтезированной речью. В работе нами использовались слоги русского языка (согласная-гласная). Согласные буквы: б, в, г, д, ж, з, к, л, м, н, п, р, с, т, ф, х (16 букв). Гласные буквы: а, у, ю, о, ё, и (6 букв). Для проверки достоверности различий использовался также однофакторный дисперсионный анализ и F-критерий. При сравнении процентов ошибок были получены следующие значения F-критерия: при сравнении синтезированной речи с «зашумленной» (соотношение С/Ш= $+20$ дб) естественной, произносимой мужским голосом $F=1.13$; при сравнении синтезированной речи с «зашумленной» (соотношение С/Ш= $+20$ дб) естественной, произносимой женским голосом $F=3.23$; при сравнении «зашумленной» естественной речи, произносимой женским голосом, с «зашумленной» естественной, произносимой мужским голосом $F=0.67$. При этом стандартное значение F-критерия для наших условий равно 4.96.

При оценке смещения слогов, сравнив полученные во всех трех случаях коэффициенты со стандартным, можно констатировать, что различия во всех трех случаях являются недостоверными.

В 14 слогах из 96 испытуемые допустили ошибки во всех трех форматах (слоги: да, ву, ку, лу, вю, гю, кю, до, по, вё, би, ви, ли, ни / 10 звонких согласных и 3 глухих). В остальных ошибки очень разнообразны. В 21 слогах совсем не было допущено ошибок (жа, ла, на, зу, ну, ру, су, фу, ху, рю, сю, фю, го, жо, зо, со, хо, жё, пё, жи, ти / 12 звонких согласных и 9 глухих). Полученные результаты также свидетельствуют, что количество ошибок в слогах со звонкими согласными незначительно превышает количество ошибок в слогах с глухими согласными.

Как уже отмечалось, перед исследователями и разработчиками стоит задача максимального приближения характеристик синтезированной речи к характеристикам естественной речи диктора. А для того, чтобы решить эту задачу,

необходимо провести оценку качества синтезированной речи. Существует несколько методов оценки качества синтезированной речи, основанных на расчете корреляции между естественным и синтезированным речевыми сигналами в пространстве различных параметров сигнала. Однако даже лучшие из них не дают результат, приближающийся к результатам субъективной оценки. Поэтому степень сходства синтезированной речи с естественной выяснялась **в экспериментах по субъективной оценке** качества речи. Методика проведения этих экспериментов основывалась на Рекомендации Р.85 ИТУ-Т «Метод субъективной оценки качества речи устройств речевого вывода» Международной комиссии по синтетической телефонии (Женева, 1994). Этот метод позволяет сравнивать нескольких систем синтеза между собой. Испытуемые выражают свое мнение в одной или более шкале оценки, отвечая на определенные вопросы по информации, которая содержится в сообщениях. Результаты - меры воспринятого качества в нескольких аспектах, которые позволяют сравнить эффективность различных систем речевого синтеза.

Используемые шкалы:

Шкала 1. «Впечатление». Шкала 2. «Усилия по слушанию». Шкала 3. «Проблемы понимания». Шкала 4. «Артикуляция». Шкала 5. «Произнесение». Шкала 6. «Нормы разговора» (предпочтительность темпа речи). Шкала 7. «Приятность голоса». Дополнительно оценивалось понимание текста (в отдельной анкете представлены вопросы по содержанию сообщения). В ходе исследования стимульный материал предъявлялся в трех форматах: речь, синтезированная TextAloud; дикторская речь; письменная речь в виде SMS сообщений.

На первом этапе испытуемым (40 человек - мужчины в возрасте от 17 до 22 лет с нормальным состоянием слуха) предлагалось прослушать простые фразы, состоящие из 2 - 4 слов, и оценить их по шкале «Общее впечатление» (при этом ответ «превосходно» соответствовал 5 баллам, «плохо» - 1 баллу).

На следующем этапе испытуемым предлагалось прослушать фразы, характерные для автоматической системы оповещения, и ответить на две анкеты, содержащие вопросы по содержанию сообщений и по оценке качества (фраза предъявлялась дважды: первый раз – для оценки содержания, второй – для оценки качества). Вопросы в анкетах соответствуют шкалам, перечисленным выше.

На третьем этапе испытуемым предлагалось прослушать типичные фразы, отправляемые в формате SMS, прочитать их и оценить (5 - максимальная оценка, 1-минимальная) качество сообщений по следующим шкалам: понятливость, предпочтение предъявляемого варианта, удобство использования, прилагаемые для понимания усилия. Эти шкалы введены нами, чтобы оценить возможное использование синтезированной речи в системах сотовой связи.

Обобщенная оценка качества речи (MOS-оценка) выражается через среднее значение оценок всех аудиторов в соответствии со шкалой. Для определения статистической значимости оценок, полученных для различных типов голосов, был осуществлён однофакторный дисперсионный анализ результатов с использованием F-критерия. Для наших условий критическое значение F будет равно 3.96. Достоверные отличия синтезированной речи от естественной получены по следующим критериям: общее впечатление от синтезированной и естественной речи ($F=22.95$), проблемы понимания (понятливость) ($F=49.0$), артикуляция ($F=60.27$), произнесение ($F=44.53$), приятность голоса ($F=50.88$), предпочтение предъявляемого варианта ($F=29.79$).

Недостовверными получились различия по следующим критериям: пониманию содержания сообщения ($F=1.95$), прилагаемым для понимания усилиям ($F=3.72$), «нормам разговора» (предпочтительности темпа речи) ($F=0.14$), удобству использования ($F= 3.25$).

Следует особо отметить, что различий нет по удобству использования того или иного формата при чтении или озвучивании сообщений SMS. Это можно объяснить тем, что удобство того или иного варианта напрямую зависит от ситуации, в которой находится пользователь.

В последнем параграфе третьей главы представлены *экспериментальные исследования понимания*. Как отмечалось ранее, для получения количественной оценки понимания повествовательных текстов при их прослушивании в нашем исследовании применяется описанная выше схема П.Торндайка. При сохранении количества пропозиций, их иерархического распределения такими, как и в рассказе «Остров круга», нами при участии магистров-психологов были разработаны три других рассказа («Гимназия», «Отдых», «Ремонт»), которые использовались в качестве эквивалентных форм при проведении исследования понимания в разных экспериментальных условиях. В исследованиях по оценке понимания приняло участие 120 человек: студенты факультета психологии университета и студенты других вузов города Екатеринбурга. Все испытуемые были разделены на три группы: группа студентов - психологов (40 человек) и две группы - студенты других вузов, сформированные по 40 человек в каждой. Исследование проводилось в два этапа. На первом этапе эксперименты во всех группах проводились перед началом занятий, когда испытуемые находились в состоянии работоспособности. Каждой группе испытуемых предъявлялся текст «Остров круга» одним из трех способов. Группе студентов - психологов текст предъявлялся с помощью системы синтеза TextAloud. Выбор именно этой группы испытуемых для прослушивания синтезированной речи объясняется тем, что они, в отличие от других групп, имели опыт работы с ней. Одна из смешанных групп получала текст посредством визуального предъявления (то есть испытуемые данной группы самостоятельно читали текст), другая – посредством прослушивания естественной речи. Задача всех участников эксперимента заключалась в том, чтобы воспроизвести данный текст как можно точнее. При обработке данных оценивались изменения количественных и качественных показателей. Качественными показателями снижения уровня понимания текста могут быть следующие: внесение побочных ассоциаций, искажение логики рассказа, подмена понятий и так далее. Что касается количественных показателей, то сравнение уровня понимания рассказов, предъявляемых разными способами, проводилось на основе разработанной нами математической модели определения коэффициента понимания (1):

$$K = \left(\sum_{i=1}^4 k_i \cdot p_i \right) \quad (1), \text{ где}$$

K – коэффициент понимания;

n – количество пропозиций в тексте на данном уровне;

p_i – количество воспроизведенных одноуровневых пропозиций

($p_{1\max}=6$; $p_{2\max}=17$; $p_{3\max}=5$; $p_{4\max}=6$);

k_i – коэффициент значимости пропозиции (максимальная значимость у пропозиций 1-го уровня; минимальная – у пропозиций 4-го уровня). С повышением уровня коэффициент значимости возрастает в геометрической

прогрессии. Данные коэффициенты являются константными: $k_1=1$; $k_2=0.5$; $k_3=0.25$; $k_4=0.125$.

Следовательно, в развернутом виде формула будет выглядеть таким образом (2):

$$K = (\sum_{i=1}^6 p_1 + 0,5 \sum_{i=1}^{17} p_2 + 0,25 \sum_{i=1}^5 p_3 + 0,125 \sum_{i=1}^6 p_4) \quad (2)$$

Для интерпретации полученных значений были рассчитаны границы диапазонов на шкале, отражающей уровень понимания текста испытуемым. Значения распределились следующим образом: 0 – 3,25 – оценивается как «плохое понимание текста»; 3,3 – 6,55 – «сниженное понимание текста»; 6,6 – 9,85 – «умеренное понимание текста»; 9,9 – 13,15 – «хорошее понимание текста»; 13,2 – 16,5 – «полное понимание текста».

В графическом виде полученные результаты представлены на рис. 1.



Рис. 1. Распределение испытуемых в соответствии с коэффициентом понимания при разных способах предъявления текста (в начале учебного дня)

Сравнение полученных результатов осуществлялось с помощью t критерия Стьюдента, посредством которого попарно сравнивались показатели понимания текста (выраженного через коэффициент понимания) в разных подгруппах испытуемых. Были получены следующие результаты: между пониманием текста, предъявляемого посредством естественной речи и пониманием текста, предъявляемого посредством синтезированной речи различия не являются статистически достоверными ($t = 1.45$ при t крит. = 3.46 для уровня статистической значимости $p \leq 0.01$); в то время как различия в понимании при визуальном и двумя другими способами предъявления текста являются статистически достоверными ($t =$

4.7 и $t = 5.2$ при сравнении с предъявлением посредством естественной и синтезированной речи соответственно).

На втором этапе исследования проводилась аналогичная процедура, но предъявлялся другой текст («Гимназия»), а главное – работа с текстом производилась в каждой подгруппе не в начале, а в конце учебного дня, когда большинство испытуемых находилось в состоянии утомления. На этом этапе экспериментально исследовалось влияние нарастающего утомления на понимание связного текста, произносимого с помощью синтезированной речи, по сравнению с пониманием текстов, предъявляемых посредством естественной речи и визуально. В графическом виде полученные результаты представлены на рис. 2.

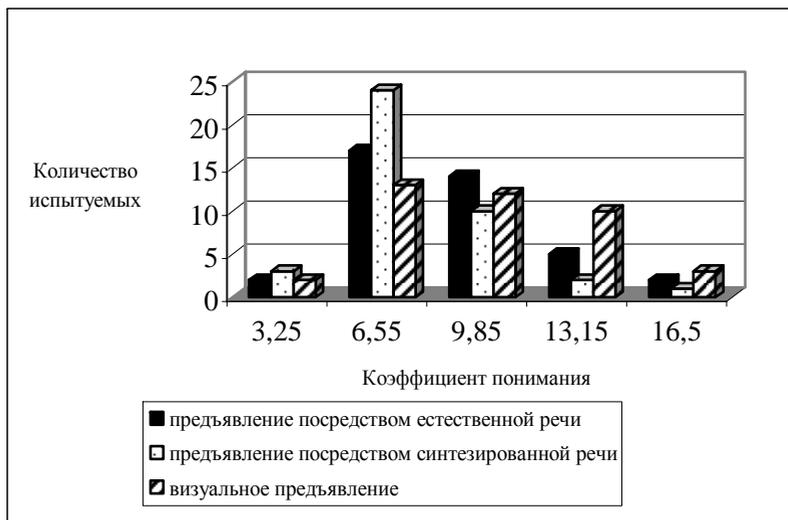


Рис.2. Распределение испытуемых в соответствии с коэффициентом понимания при разных способах предъявления текста (в конце учебного дня)

Также, как и на первом этапе исследования, здесь был рассчитан критерий Стьюдента. Полученные результаты свидетельствуют: между пониманием текста, предъявляемого посредством естественной речи и пониманием текста, предъявляемого посредством синтезированной речи, различия являются статистически достоверными ($t = 4,1$ при t крит. = 3,46 для уровня статистической значимости $p \leq 0,01$); различия в понимании при визуальном и двумя другими способами предъявления текста также являются статистически достоверными ($t = 5,2$ и $t = 5,9$ при сравнении с предъявлением посредством естественной и синтезированной речи соответственно).

Следовательно, в состоянии утомления появляются статистически значимые различия между пониманием синтезированной и естественной речи. Этот факт можно объяснить с позиции снижения общего запаса ресурсов субъекта, находящегося в состоянии утомления. По сравнению с состоянием

работоспособности, у него появляются более выраженные затруднения в дешифровке текста, предъявляемого посредством синтезированной речи. Это происходит из-за того, что субъект затрачивает немалые ресурсы в преодолении трудностей восприятия такой речи, а оставшихся ресурсов, в условиях их ограниченности, оказывается недостаточно для достижения высокого уровня понимания текста. Однако различия являются преодолимыми, и, по всей видимости, путем дальнейшей тренировки могут быть нивелированы.

Что касается визуально предъявленного текста, то уровень его понимания остается более высоким. Эту закономерность можно объяснить тем, что школьники и студенты чаще имеют дело с визуальным способом предъявления текстов. Кроме этого, надо учитывать процессуальные особенности чтения, отличающие его от слушания как естественной, так и синтезированной речи. В процессе чтения текста субъект может неоднократно возвращаться к нему или к отдельным его единицам, в процессе слушания возвращение к тексту затруднено.

По окончании работы над основными сериями были проведены дополнительные серии, которые проходили по той же схеме: три группы испытуемых проходили тестирование сначала в состоянии работоспособности, а затем в состоянии утомления. Участвовали те же группы испытуемых. В качестве стимульного материала выступали тексты «Отдых» и «Ремонт». Дополнительные серии были проведены для определения ретестовой надежности разработанного метода.

Ретестовая надежность была рассчитана с помощью коэффициента корреляции Пирсона. Парно сравнивались первоначальные показатели коэффициента понимания в каждой из трех групп испытуемых с аналогичными ретестовыми показателями (отдельно для состояния работоспособности и состояния утомления). Результаты представлены в таблице 3.

В целом, в соответствии с полученными результатами можно заключить следующее.

1. Понимание синтезированной речи сопоставимо с пониманием естественной речи, в то время как оба они отличаются от понимания визуально предъявленного текста, которое достигает более высокого уровня у большинства испытуемых.

2. Разработанный нами метод количественной оценки понимания синтезированной речи показал высокую надежность результатов оценки понимания повествовательных текстов при их прослушивании, что в свою очередь дает основание для доверия полученным с его помощью данным. Существующие незначительные различия в результатах первоначальных и ретестовых серий можно объяснить тем, что несмотря на единую пропозициональную структуру предъявляемых текстов, содержание в них отличается.

Таблица 3

Показатели ретестовой надежности разработанного метода оценки понимания

	Естественная Речь		Синтезированная Речь		визуальное предъявление	
	r_x	Уровень значимости	r_x	Уровень значимости	r_x	Уровень значимости

	$r_{x \text{ крит}} = 0,39$ (для $p < 0,01$)					
В начале дня	0,82	$p < 0,01$	0,78	$p < 0,01$	0,83	$p < 0,01$
В конце дня	0,81	$p < 0,01$	0,76	$p < 0,01$	0,81	$p < 0,01$

Выводы по работе

1. На правильность распознавания речи слушателем влияет ряд факторов: 1) конкретные требования и условия, связанные со спецификой решаемой задачи; 2) ограничения, присущие системе обработки информации, которой наделен человек; 3) опыт и тренировка слушателя; 4) лингвистическая структура сообщения и 5) структура и качество речевого сигнала.

2. Правильность распознавания русскоязычной речи, синтезированной с помощью любой системы «текст - речь» при ограниченном числе возможных ответов, была ниже, чем естественной речи даже при отсутствии ограничений на набор ответов.

3. Совокупность полученных результатов экспериментальных исследований позволяет утверждать, что перцептивные процессы по расшифровке синтезированной речи требуют больших умственных «усилий» или ресурсов, чем процессы расшифровки естественной речи.

4. Главный вывод на основе полученных результатов при изучении смещения слогов и стробирования сигнала заключается в том, что в отличие от экспериментов американских авторов не выявлено заметной разницы в восприятии акустико-фонетической структуры русскоязычной синтезированной речи по сравнению с естественной. В наших экспериментах со стробированием сигнала показано, что средняя длительность слова, необходимая для правильной идентификации, как для естественной, так и для синтезированной речи составляет 65% всего слова ($p < 0,05$). В экспериментах со смещением слогов не выявлено достоверных отличий синтезированной речи в сравнении с зашумленной естественной. Отсюда следует, что синтезированная речь по параметрам близка к естественной речи, однако в отличие от нее является менее четко выраженной.

5. Понимание синтезированной речи отличается от понимания естественной речи в состоянии утомления. Здесь появляются статистически значимые различия между пониманием синтезированной и естественной речи, что объясняется снижением общего запаса когнитивных ресурсов человека, находящегося в состоянии утомления. По сравнению с состоянием работоспособности у слушателя появляются более выраженные затруднения в дешифровке текста, предъявляемого посредством синтезированной речи. Он вкладывает значительные ресурсы в преодоление трудностей восприятия такой речи, а оставшихся, ввиду их ограниченности, оказывается недостаточно для достижения высокого уровня понимания текста.

6. На основе теоретического анализа отечественных и зарубежных моделей понимания нами определено, что наиболее перспективны с точки зрения количественного оценивания понимания повествовательных текстов,

воспроизводимых синтезированной или естественной речью, основанные на пропозициях модели У. Кинча и П. Торндайка.

7. Разработанная нами математическая модель, основанная на теории П. Торндайка, доказала свою эффективность и надежность при количественном оценивании понимания синтезированной речи.

**Список работ,
опубликованных по теме диссертации:**

Статья, опубликованная в ведущем рецензируемом научном журнале, рекомендованном ВАК:

1. Санникова Е.Г., Воронин В.М., Ильиных Т.Б. «Психологические проблемы эффективного функционирования системы врач – современные информационные технологии». // Проблемы управления здравоохранением. – Москва, 2006, № 5. С. 39 – 43.

Брошюры и статьи в сборниках:

2. Санникова Е.Г. Психологические проблемы восприятия естественной и синтезированной речи. Екатеринбург: Изд-во СОБСС, 2007, 29 стр.

3. Санникова Е.Г., Воронин В.М. Психологические проблемы понимания естественной и синтезированной речи. Екатеринбург: Изд-во СОБСС, 2007, 29 стр.

4. Санникова Е.Г., Воронин В.М., Быкова Н.И., Миронова А.Р. Новый подход к оценке ментального утомления. // Психологический вестник Уральского гос. ун-та им. А.М. Горького. – Екатеринбург: Изд-во Уральского университета, 2008. С. 194-198.

5. Санникова Е.Г., Воронин В.М., Быкова Н.И., Миронова А.Р. Исследование понимания текстов, предъявляемых в визуальной и аудиальной форме в состоянии работоспособности и утомления. // Психологическое обеспечение качества образования в условиях его модернизации: материалы второй Всероссийской научно-практической интернет – конференции. Изд-во Белгородского университета, 2008. С. 35-46.

