ИНФОРМАТИКА. ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА. УПРАВЛЕНИЕ

УДК 004.896
DOI: 10.35330/1991-6639-2019-3-89-3-14

БАЗОВЫЕ ЭЛЕМЕНТЫ КОГНИТИВНОЙ МОДЕЛИ МЕХАНИЗМА ВОСПРИЯТИЯ РЕЧИ НА ОСНОВЕ МУЛЬТИАГЕНТНОГО РЕКУРСИВНОГО ИНТЕЛЛЕКТА*

З.В. НАГОЕВ[1], И.А. ГУРТУЕВА[2]

[1]ФГБНУ «Федеральный научный центр
«Кабардино-Балкарский научный центр Российской академии наук»
360002, КБР, г. Нальчик, ул. Балкарова, 2
E-mail: cgrkbncran@bk.ru
[2]Институт информатики и проблем регионального управления –
филиал ФГБНУ «Федеральный научный центр
«Кабардино-Балкарский научный центр Российской академии наук»
360000, КБР, г. Нальчик, ул. И. Арманд, 37-а
E-mail: iipru@rambler.ru

В данной работе проанализирована обобщенная архитектура, лежащая в основе практически всех современных систем автоматического распознавания речи. Кратко изложена необходимость разработки принципиально нового подхода к решению проблем распознавания речи. Предлагается формальное описание структуры акта речевосприятия для применения в качестве общей теоретической основы при разработке универсальных систем автоматического распознавания речи, высокоэффективных в условиях высокой зашумленности и ситуациях «cocktail party». Разработана общая структурная динамика процесса распознавания речи, позволяющая учесть лингвистические и экстралингвистические аспекты речевого сообщения. Доказана необходимость использования понятия артикуляционного события в качестве минимального базового паттерна распознавания звукового образа.

Процесс распознавания структурирован на основе функциональной детерминанты «ситуация». Необходимость анализа многочисленных источников информации, сопровождающих звуковое сообщение, отказ от поиска инварианта носят здесь принципиальный характер.

Формальными средствами для реализации выбраны мультиагентные системы. Мультиагентный подход позволяет дифференцировать и анализировать звуки разной природы. Это делает предложенную модель уникальной и дает ей преимущества в ситуации так называемой «cocktail party», а также в задачах, где уровень шумов крайне высок.

Ключевые слова: *мультиагентные системы, искусственный интеллект, искусственные нейронные сети, распознавание речи.*

1. Введение

Бурное развитие систем обволакивающего интеллекта предъявляет все более строгие требования к речевым системам. Основные проблемы выявляются при распознавании речи в зашумленных условиях, при анализе акустических сцен (так называемой ситуации «cocktail party») [16, 34]. Все еще актуально решение проблемы акустической вариативности [16, 27, 33]. Системы автоматического распознавания речи, предлагаемые современными IT-технологиями, не способны решить указанные задачи с удовлетворительной точностью и не могут быть признаны универсальными [10, 27]. На наш взгляд, успешное решение данной проблемы возможно как следствие решения проблем искусственного интеллекта, поскольку необходимо использование внутренней семантической модели, построенной на основе когнитивных функций, которыми

* Работа выполнена при поддержке грантов РФФИ №№ 18-01-00658, 19-01-00648

пользуется человек при декодировании звуковых сообщений. В работе [20] предложен подход к формализации семантики разумного мышления с использованием когнитивного моделирования на основе концепции рекурсивной когнитивной архитектуры и гипотезы об инварианте организационно-функциональной структуры процесса интеллектуального принятия решения на основе когнитивных функций.

В настоящей статье приведен краткий обзор методов и алгоритмов, наиболее широко применяемых для решения проблем автоматического распознавания речи, и предложены основные элементы когнитивной модели распознавания речи как теоретической базы для разработки нового подхода к исследованию указанных задач.

Актуальность исследования состоит в том, что отсутствие надежных систем распознавания речи во многом является сдерживающим фактором развития интеллектуальной робототехники. Объектом исследования данной работы является распознавание речи в реальных средах на основе когнитивного моделирования. В качестве предмета исследования рассматривается структурно-функциональная организация слухового анализатора интеллектуального агента на основе нейрокогнитивных архитектур.

2. Структура процесса автоматического распознавания речи

Задача распознавания речи, в наиболее общем понимании, сводится к проблеме многоуровневого распознавания образов, при решении которой акустические сигналы анализируются и структурируются в иерархию элементов слова (фонем, морфем или слогов), слов, словосочетаний и предложений [14, 16]. Каждый уровень имеет дополнительные временные пределы, например, известные произношения слов или разрешенные последовательности слов, которые могут компенсировать ошибки и неопределенности на нижних уровнях. Иерархия ограничений эксплуатируется наилучшим образом комбинацией принятия решений на основе вероятностных моделей на нижних уровнях и принятия дискретных решений только на высоких уровнях.

Современные речевые приложения проектируются в соответствии с конкретными целями и задачами и потому могут использовать самые разные алгоритмы и методы, но обобщенный процесс автоматического распознавания речи можно разделить на этапы, показанные на рис. 1.

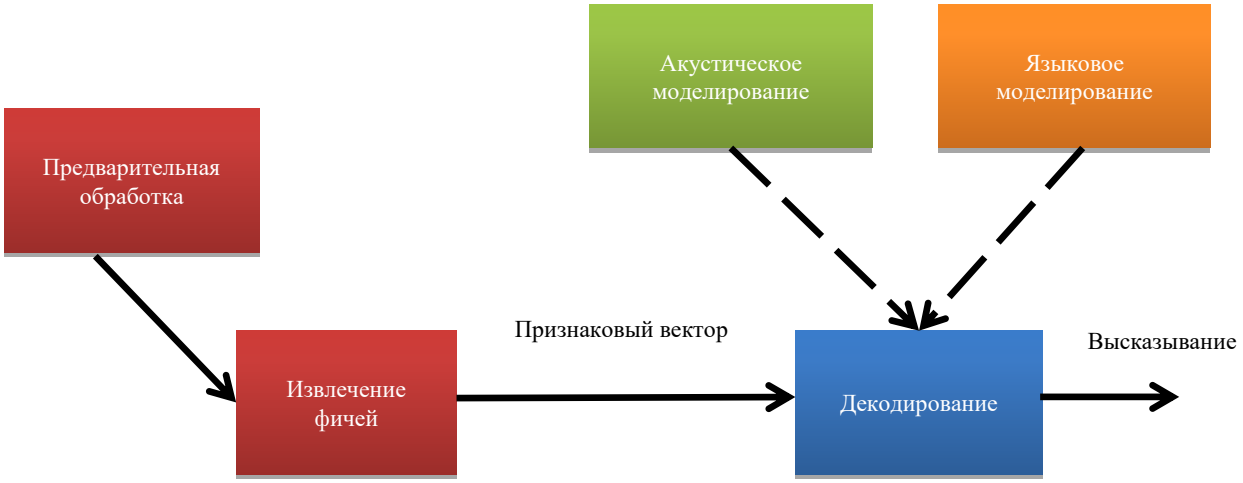

Рис. 1. *Обобщенная архитектура автоматических систем распознавания речи*

Подсистема предварительной обработки сигнала обрабатывает сигнал для исключения влияния окружающей среды, улучшения качества и выделения полезного сигнала [14, 26]. Методы, применяемые на данном этапе, можно разделить на несколько групп.

Для шумоочистки исследуемого сигнала применяются, как правило, три подхода: обучение в условиях высокой зашумленности, модификация эталонов с использованием оценки уровня шумов, а также выделение инвариантных признаков относительно шума. Наиболее популярными методами первого подхода являются коэффициенты линейного предсказания и кепстральные коэффициенты [26]. Модификация эталонов связана с цифровой обработкой сигналов. Наиболее популярны методы маскирования шумов и методы шумоподавления с использованием нескольких микрофонов. Последний подход чаще всего использует кратковременную функцию когерентности [13] или модели слуховой системы человека [9].

Пространственная обработка осуществляется методами спектрально-пространственной фильтрации многомерного звукового сигнала [28]. В большинстве подходов используются спектральные методы обработки звукового сигнала, а также параметрические методы для статистической оценки положения источника звука и фильтрации полезного сигнала.

Среди спектральных подходов удобны в использовании методы формирования луча [30], сводящиеся к выделению компонент сигнала, распространяющегося из определенной точки пространства. Простейшим из них является метод «суммирования и задержки» [28], основанный на взвешенном суммировании сигналов с задержкой. Основной недостаток всех предложенных методов наблюдается в условиях высокой реверберации. Для повышения точности локализации эффективна интеграция методов аудио- и видеообработки [16].

Кроме того, в отдельную группу можно отнести методы повышения надежности распознавания речи с использованием неречевой информации о дикторе. Например, информация об артикуляции диктора, о его местоположении, об эмоциональном состоянии диктора, регистрация выдыхаемого воздуха [14].

Затем на этапе извлечения фичей входной речевой сигнал преобразуется в набор акустических параметров. Речь трансформируется для выявления релевантных характеристик и сжимается для упрощения последующей обработки. Методы параметризации, применяемые на данном этапе, можно разделить на три группы: анализ, основанный на производстве речи, анализ восприятия и методы, описывающие сигнал с точки зрения его спектральных составляющих [14].

Методы спектрального анализа не рассматривают, как происходит порождение и восприятие речи. Единственное их исходное предположение в том, что речь стационарна. Быстрое преобразование Фурье (БПФ) [26] – самый широко применяемый алгоритм для извлечения фичей. БПФ осуществляет переход от амплитуд к дискретным частотам по времени и интерпретируется визуально.

Методы, разработанные на основе анализа процесса речепроизводства, описывают речевой сигнал как комбинацию источника звуковой энергии и функции переноса (фильтра), которая ее модулирует. Функция переноса определяется формой речевого тракта и может моделироваться как линейный фильтр. Источник можно классифицировать по двум типам. Первый – квазипериодический, который имеет место при открытии связок (производство звонких звуков). Этот источник можно моделировать как последовательность импульсов. Второй связан с возбуждением без участия голоса. В данном типе голосовые связки разведены, но некоторые ограничения осложняют свободное движение воздушного потока. Этот источник можно моделировать как случайный сигнал. Данная модель не объясняет производство звонких фрикативов. Звонкие фрикативы произносятся как смесь источников возбуждения: периодичного компонента и аспирации. Такая комбинация не учитывается моделью «источник-фильтр».

Наиболее применяемыми методами анализа речепроизводства являются спектральный конверт, кепстральный анализ и анализ линейного предсказания [16, 28, 31]. Спектральный конверт может быть использован для выделения речевых единиц, которые

лингвистически различны в данном языке. Следовательно, целью многих техник речевого анализа является отделение спектрального конверта (формы фильтра) от источника. Кепстральный анализ осуществляет деконволюцию источника возбуждения и ответ речевого тракта. Обычно эта операция невозможна для звуковых сигналов в общем, но для речевых сигналов это возможно, поскольку оба сигнала имеют различные спектральные характеристики [28]. Линейное предсказание предполагает, что выход акустического фильтра можно аппроксимировать линейной комбинацией прошлых речевых образцов и некоторого входящего возбуждения. Основной подход сводится к отысканию набора коэффициентов предсказания, которые минимизируют среднеквадратичную ошибку предсказания речевого сегмента. С учетом того, что спектральные характеристики фильтра речевого тракта зависят от времени, коэффициенты предсказания оцениваются в короткие сегменты времени (кратковременный анализ).

Анализ восприятия для представления речевого сигнала использует некоторые аспекты и поведение слуховой системы человека. Наиболее успешно применяются два метода: линейное предсказание восприятия и метод кепстральных коэффициентов [16, 28].

Метод кепстральных коэффициентов строит речевое представление на основе свойства нелинейности слуховой системы, деформирует линейный спектр в нелинейную частотную Mel-шкалу. Mel-шкала моделирует чувствительность человеческого уха. Для распознавания речи обычно используют первые 13 коэффициентов кепстра [14]. Одним из «плюсов» данного метода является то, что он устойчив к искажениям канала свертки [14, 28]. Перцептивное линейное предсказание модифицирует короткопериодный спектр речи с помощью спектральных трансформаций, основанных на данных психофизиологических исследований, предшествующих анализу линейного предсказания [28].

Далее речевой сигнал фреймируется и осуществляется акустический анализ [8-10, 31]. Существует большое число акустических моделей, отличных друг от друга по их представлению, зернистости, контекстной зависимости и другим свойствам. Акустический анализ осуществляется наложением каждой акустической модели на каждый речевой фрейм; на выходе получается матрица оценки фреймов. Оценки вычисляются в зависимости от типа использованной акустической модели. Для акустических моделей, основанных на шаблонном подходе, оценки представляют собой Евклидово расстояние между фреймом шаблона и фреймом неизвестного сигнала. Для акустических моделей на состояниях оценка представляет собой эмиссионную вероятность, то есть сходство текущего состояния, генерирующего текущий фрейм согласно параметрической или непараметрической функции состояния.

Шаблонный подход эффективен для систем с малыми словарями, которые могут поддерживать модели на основе целого слова. В крупных системах используется более гибкое представление, основанное на обучаемых акустических моделях или состояниях. Наиболее эффективное акустическое моделирование основано на структуре, называемой скрытое Марковское моделирование [14].

Оценка фреймов конвертируется в последовательность слов на основе идентификации последовательности акустических моделей, представляя собой значимую последовательность слов, которая дает наилучшую суммарную оценку вдоль пути выравнивания по матрице. Процесс поиска наилучшего пути выравнивания называется временем выравнивания. Выравнивание по времени эффективно осуществляется с помощью динамического программирования, общего алгоритма, который использует ограничения по локальному пути и имеет требования к линейности по времени и пространству. Общий алгоритм имеет два основных варианта, известных как динамическое искажение времени [14] и поиск Витерби [14], несколько отличающихся в локальных вычислениях и критериях оптимальности. Конечным результатом

выравнивания по времени является последовательность слов как гипотеза о предложении, соответствующем распознаваемому высказыванию. На практике обычно возвращается несколько таких предложений с наивысшими оценками с использованием варианта выравнивания по времени, называемого N-best search [31]. Это позволяет системе распознавания проложить два пути вдоль неизвестного высказывания: первый – на основе упрощенных моделей для быстрой генерации наилучшего списка вариантов (N-best list), второй – более сложных моделей для точной переоценки каждой из N-гипотез и возвратить наилучшую из них.

В настоящее время на рынке существует достаточное число речевых приложений с высокой точностью распознавания, но их эффективность оценивается в узких условиях, при переносе из лабораторных в реальные условия эксплуатации существующие системы недостаточно устойчивы к шумам и бесполезны в случае “cocktail party”, а также в условиях конференции или совещания. Существует необходимость разработки принципиально нового подхода, базирующегося на иных математических методах, для решения проблем распознавания речи.

На наш взгляд, успешное решение данной проблемы возможно как следствие решения проблем создания искусственного интеллекта, поскольку необходимо использование внутренней семантической модели, построенной на основе когнитивных функций, которыми пользуется человек при декодировании звуковых сообщений. В работе [20] предложен подход к формализации семантики разумного мышления с использованием когнитивного моделирования на основе концепции рекурсивной когнитивной архитектуры и гипотезы об инварианте организационно-функциональной структуры процесса интеллектуального принятия решения на основе когнитивных функций. Эти разработки, как и предложенный в [20] метод обучения мультиагентных нейроподобных систем на основе онтонейроморфогенеза, направлены на создание самоорганизующихся мультиагентных эмерджентных систем, способных к эмуляции когнитивных функций, используемых человеком при распознавании речи. Это позволит построить систему распознавания речи в среде с несколькими дикторами на основе имитации механизма внимания, то есть создать систему распознавания речи, автоматически фокусирующуюся на конкретном дикторе или интересующей теме.

3. Модель когнитивного механизма распознавания речи на основе мультиагентных систем

Всякое речевое сообщение обладает информационной избыточностью [33] и может быть исследовано с позиций разных научных дисциплин [5]. Результаты психоакустических исследований подтверждают тот факт, что человек при декодировании сигнала активно использует вспомогательную неречевую информацию. По этой причине в последнее время для повышения надежности распознавания или коррекции результатов при проектировании речевых систем используются видеокамеры и фотодатчики для анализа информации об артикуляции, датчики, регистрирующие электрическую активность кожи, для ввода информации об эмоционально-личностном состоянии диктора и т.д. [16]. Таким образом, для успешного решения задачи распознавания речи необходимо анализировать речь в единстве всех указанных аспектов.

Когнитивное моделирование рассматривает распознавание речи как задачу многоуровневого распознавания образов. Фундаментальные основы данных разработок и метод обучения мультиагентных нейроподобных систем на основе онтонейроморфогенеза детально описаны в работах [20-24]. Данный подход опирается на теоретический фундамент когнитивной психологии и когнитивной нейрологии [2, 7, 11, 17, 25], а также на современные достижения информатики [1, 4, 6, 18, 29], в частности мультиагентный подход [15, 32].

На рисунке 2 изображена принципиальная структура динамики акта речевосприятия в виде когнитивной архитектуры. Акт речевосприятия как устойчивая фазовая последовательность, регулярно воспроизводимая индивидуумом в речевой деятельности, может быть структурирован в виде поступательной обработки звуковой информации на следующих уровнях: предварительное распознавание (I уровень), подсознательное распознавание (II уровень), осознанное распознавание (III уровень), уровень ситуаций (IV уровень).

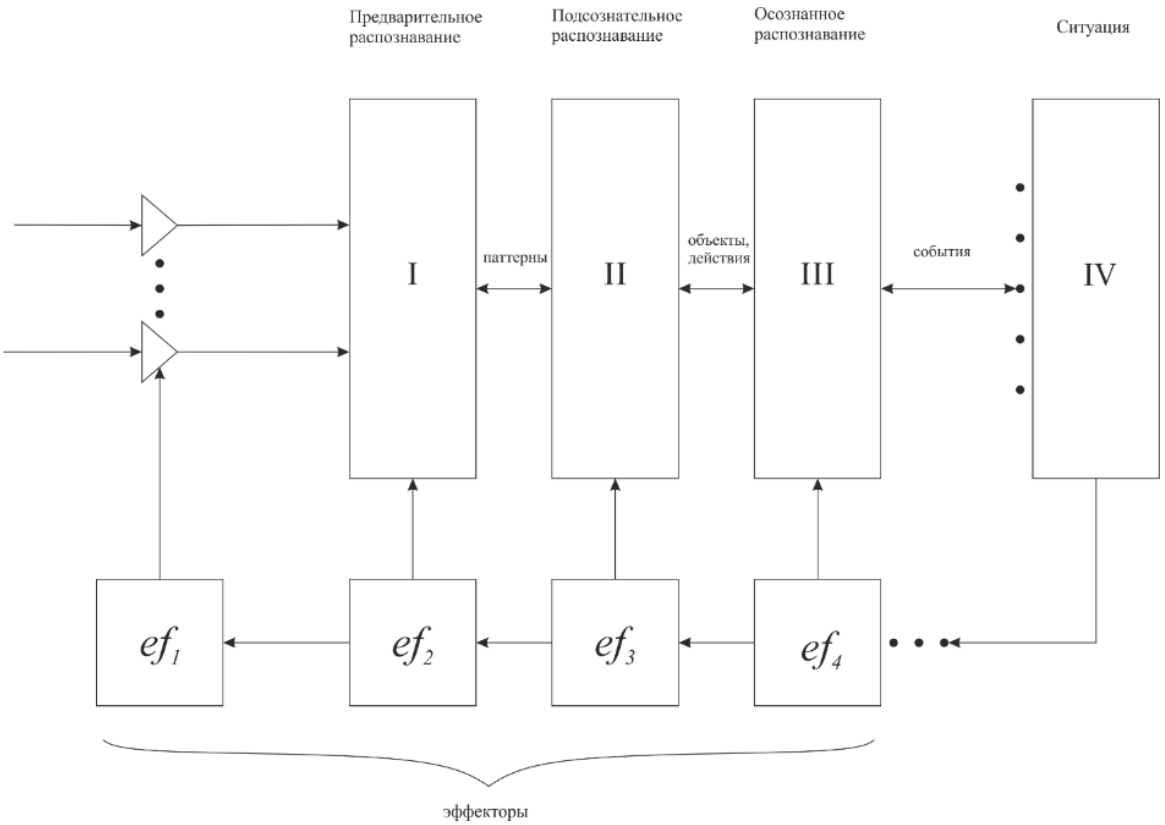

Рис. 2. *Когнитивная архитектура системы автоматического распознавания речи на основе мультиагентных моделей*

Первый уровень архитектуры – «предварительное распознавание» – имитирует регистрацию акустических параметров сигнала слуховыми рецепторами человека. Поскольку звуковую волну в основном можно охарактеризовать четырьмя физическими параметрами, мы считаем необходимым выделить в структуре информационного потока афферентного слухового тракта четыре слоя, а именно: амплитуду сигнала, частоту звуковой волны, продолжительность звучания и, наконец, местоположение источника сигнала. Каждая акустическая составляющая вызывает определенное слуховое ощущение, описываемое физиологическим параметром: частоте звукового стимула соответствует высота звука, интенсивности (или амплитуде) – громкость. Длительность характеризуется одной величиной и в физиологии, и в акустике. Важной составляющей в информационном потоке являются сведения о пространственной локализации источника звукового сообщения.

На данном этапе обработки сигнал преобразуется в набор сигнатур, на основе которого создается матрица, в полном объеме характеризующая акустические особенности сигнала [24]. Затем на основе алгоритма машинного обучения, аналогичного подробно описанному в работе [23], осуществляется формирование множеств агентов, соответствующих каждой

минимальной речевой единице языка. Данный алгоритм позволяет снизить остроту проблемы высокой вариативности речи на базе классификации [3]. Важно отметить, что на данном уровне впервые осуществляется попытка семантизации, установления смысловой связи между обозначающим (фонемой) и обозначаемым (литерой).

На следующем уровне – «подсознательное распознавание» – сигнатуры предыдущего слоя группируются вокруг значимых объектов и действий. Мы полагаем, что для надежного распознавания речи необходимо установить связь спектральных характеристик сигнала и «артикуляционного события», лежащего в его основе. Артикуляционным событием мы называем пару агентов, один из которых идентифицирует объект, а другой – действие, а также их контракт, то есть динамическую связь, послужившую источником звука. На этапе обучения эксперт комментирует звуковые сигналы на естественном языке следующим образом: «Машина проехала», «Иван Петрович сказал» и т.п.

Если рассматривать артикуляционное событие в терминах лингвистики, можно сказать, что один агент представляет собой тему высказывания, а второй – рему. Мы опираемся на актуальное членение речевого предложения, поскольку в отличие от формального членения, отражающего грамматический аспект сообщения, актуальное синонимично логико-смысловому. Актуальное членение высказывания рассматривает сообщение с коммуникативной и когнитивной точек зрения. Интересно отметить, что немецкие младограмматики называли тему и рему психологическими субъектом и предикатом. И хотя в лингвистике принято считать данные термины неудачными, возможно, в психолингвистике стоит их использовать, поскольку они удачно иллюстрируют тот факт, что сочетание психологического субъекта и предиката формирует первичную основу контекста и становится ключом для «понимания» высказывания на следующих уровнях обработки.

Выбор артикуляционного события, а также его связи со спектральными характеристиками высказывания в качестве самостоятельного объекта анализа хорошо согласуется с основным постулатом моторной теории восприятия речи о том, что невербальную звуковую информацию человек идентифицирует как некоторое событие. Например, интерпретируя неречевой сигнал, человек использует простые высказывания: «послышались шаги», «пропела птица» [19]. Аналогичная идентификация вербальных событий становится основой для формирования механизма направленного внимания и решения проблемы контекста.

Наконец, нейропсихологи выделяют в структуре слуховой системы две подсистемы: неречевой и речевой слух [19]. По результатам клинических исследований известно, что у правшей способность дифференцировать и анализировать вербальные сообщения (фонематический или речевой слух) нарушается при поражении левой височной доли коры головного мозга, а невербальные – правой. В наиболее тяжелых случаях слуховой агнозии пациенты не могут определить смысл простейших бытовых звуков (скрипа дверей, например).

Речевой слух, в свою очередь, неоднороден и включает в себя фонематический и интонационный слух. Адекватная обработка интонационного содержания речи так же, как и невербальных сообщений, нарушается в случаях правосторонней локализации поражения (у правшей). Относительная автономность артикуляционного события как самостоятельного объекта анализа дает возможность имитировать дифференциацию обработки вербальных и невербальных звуковых образов разными полушариями мозга.

Таким образом, с точки зрения нейропсихологии введение такого понятия, как артикуляционное событие, необходимо для построения корректного механизма речевосприятия. Артикуляционное событие позволяет внести в рамки анализа большой объем информации, передаваемой интонационными средствами (эмоциональное содержание высказывания, модальность и даже нормы языка), формализовать

возникновение и построение контекста, включить в сферу анализа коммуникативные намерения, а также соответствующие психологические и поведенческие реакции.

На третьем уровне – уровне «осознанного распознавания» – выделяются значимые по текущему приоритету события, определяемые на основе работы т.н. когнитона эмоциональной оценки [20] – функционального узла мультиагентного рекурсивного интеллекта, содержащего априорную и приобретенную на основе обучения информацию о степени значимости событий для реализации целевой функции интеллектуального агента. Агенты следующих уровней мультиагентного интеллекта [20] – когнитонов целеполагания и синтеза планов действий – формируют управляющие команды для эффекторов тонкой настройки системы фильтрации и усиления акустических параметров интегрирования в афферентный тракт.

На четвертом уровне формируется «ситуация» [20], звуковой элемент увязывается с общим контекстом, включая экстралингвистические связи, и строятся прогнозы.

Таким образом, предлагаемая модель когнитивного механизма восприятия речи позволяет включить в процедуру анализа сигнала все аспекты речевого сообщения, включая экстралингвистическую составляющую, выражаемую в данном подходе в терминах события и ситуации. Каждый последующий уровень структурирования сигнала в иерархию элементов слова, слов, фраз и т.д. имеет дополнительные временные пределы, например, известные произношения слов или разрешенные последовательности слов, которые могут компенсировать ошибки и неопределенности на нижних уровнях. Иерархия ограничений применяется для организации взаимодействия между уровнем принятия решений и на основе контрактных отношений.

4. Заключение

Разработаны элементы когнитивной модели распознавания речи на основе мультиагентной рекурсивной когнитивной архитектуры, позволяющей учесть лингвистические и экстралингвистические составляющие речевого сообщения. В качестве базового паттерна распознавания звукового образа выбрано артикуляционное событие. Процесс распознавания структурирован на основе функциональной детерминанты «ситуация». Необходимость анализа многочисленных источников информации, сопровождающих звуковое сообщение, отказ от поиска инварианта носят здесь принципиальный характер.

За счет мультиагентной природы, использования пространственно-временных характеристик и самообучения данный подход позволяет отделить друг от друга и проанализировать звуки разной природы.

ЛИТЕРАТУРА

1. *Abdel-Hamid O., Mohamed A., Jiang H., Penn G.* Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition, Proc. IEEE Int. Conf. Acoust., Speech, Signal Process, 2012. Pp. 4277-4280.

2. *Chomsky N.A.* A Review of Skinner's Verbal Behavior, [Readings in the Psychology of Language], Prentice-Hall, Upper Saddle River, New Jersey, 1967. Pp. 636.

3. *Coates A., Ng A.Y.* Learning feature representations with *k*-means, Neural Networks: Tricks of the Trade, 2012. Pp. 561-580.

4. *De Mulder W., Bethard S., Moens M.-F.* A Survey on the Application of Recurrent Neural Networks to Statistical Language Modeling, Computer Speech and Language, 2015. № 30(1). Pp. 61-98.

5. *Де Соссюр Ф.* Курс общей лингвистики. Екатеринбург: Изд-во Уральского университета, 1999. Pp.256.

6. *Deng L., Li X.* Machine Learning Paradigms for Speech Recognition: An Overview, IEEE Transactions on Audio, Speech, and Language Processing, 2013. № 21(5). Pp. 1060-1089.

7. *Gazzaniga M.S.* Conversations in the Cognitive Neuroscience. The MIT Press, Cambridge, 1996. Pp. 752.

8. *Ghai W., Singh N.* Literature Review on Automatic Speech Recognition // International Journal of Computer Applications. 2012. № 41 (8). Pp. 42-50.

9. *Ghitza O.* Auditory nerve representation as a front-end for speech recognition in a noisy environment // Computer Speech and Language, 1986. Vol. 1. Pp. 109-130.

10. *Gupta V.* A Survey of Natural Language Processing Techniques // International Journal of Computer Science & Engineering Technology. 2014. № 5 (1). Pp. 14-16.

11. *Haikonen P.* The Cognitive Approach to Conscious Machines, imprint Academic, Exeter, UK, 2003. Pp. 300.

12. *Hinton G., Deng L., Yu D., et al.* Deep neural networks for acoustic modeling in speech recognition // IEEE Signal Process. Mag. 2012. № 29(6). Pp. 82-97.

13. *Juan B.H.* Speech Recognition in Adverse Environments. Computer Speech and Language, 1991. Vol. 5. Pp. 275-294.

14. *Jurafsky D., Martin J.* Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition. Prentice Hall, Boston, 2008. Pp. 1032.

15. *Kotseruba I., Tsotsos J.K.* A Review of 40 Years of Cognitive Architecture Research: Core Cognitive Abilities and Practical Applications, 2016 available at: arxiv.org/abs/1610.08602.

16. *Мазуренко И.Л.* Компьютерные системы распознавания речи // Интеллектуальные системы. 1998. Т. 3. Вып. 1-2. С. 117-134.

17. *Minsky M.* The Society of Mind. Simon and Shuster. New York, 1988. Pp. 336.

18. *Mohamed A., Dahl G., Hinton G.* Acoustic modeling using deep belief networks // IEEE Audio, Speech, Lang. Process. 2012. № 20(1). Pp. 14-22.

19. *Морозов В.П., Вартанян И.А., Галунов В.И.* Восприятие речи: вопросы функциональной асимметрии мозга. Ленинград: Наука, 1988, 135 с.

20. *Нагоев З.В.* Интеллектика, или Мышление в живых и искусственных системах. Нальчик: Изд-во КБНЦ РАН, 2013. С. 232.

21. *Нагоев З.В., Нагоева О.В.* Извлечение знаний из многомодальных потоков неструктурированных данных на основе самоорганизации мультиагентной когнитивной архитектуры мобильного робота // Известия КБНЦ РАН. 2015. № 6 (68). С. 73-85.

22. *Нагоев З.В., Нагоева О.В.* Зрительный анализатор интеллектуального робота для обработки неструктурированных данных на основе мультиагентной нейрокогнитивной архитектуры // Перспективные системы и задачи управления: материалы XII Всероссийской научно-практической конференции (г. Ростов-на-Дону, 2017 г.). ЮФУ, 2017. С. 457-467.

23. *Нагоев З.В., Денисенко В.А., Лютикова Л.А.* Система обучения автономного сельскохозяйственного робота распознаванию статических изображений на основе мультиагентных когнитивных архитектур // Устойчивое развитие горных территорий. 2018. № 2. С. 289-297.

24. *Nagoev Z., Lyutikova L., Gurtueva I.* Model for Automatic Speech Recognition Using Multi-Agent Recursive Cognitive Architecture, Annual International Conference on Biologically Inspired Cognitive Architectures BICA, 2018, Prague, Chech Republic http://doi.org/10.1016/j.procs.2018.11.089

25. *Newell A.* Unified Theories of Cognition. Harvard University Press, Cambridge, Massachusetts, 1990. Pp. 576.

26. *Rabiner L.R., Schafer R.W.* Digital Processing in Speech Signal. Moscow: Radio and communications, 1981. Pp. 496.

27. *Reddy R.* Speech Recognition by Machine: A Review, Proceedings of the IEEE, 1976. № 64 (4). Pp. 501-531.

28. *Ронжин А.Л., Карпов А.А., Кагиров И.А.* Особенности дистанционной записи и обработки речи в автоматах самообслуживания // Информация и системы управления. 2009. № 5. Pp. 32-38.

29. *Schunk D.H.* Learning Theories: An Educational Perspective, Pearson Merrill Prentice Hall, Boston, 2011. Pp. 576.

30. *Van Veen B.D., Buckley K.M.* Beamforming: A Versatile Approach to Spatial Filtering // IEEE ASSP Magazine. 1988. № 5(2). Pp. 4-24.

31. *Waibel A., Lee K.-F.* Readings in Speech Recognition, Morgan Kaufman, Berlington, 1990. Pp. 680.

32. *Wooldridge M.* An Introduction to Multi-Agent Systems, Wiley, Hoboken, 2009. Pp. 366.

33. *Зиндер Л.Р.* Общая фонетика, М.: Высш. школа, 1979. С. 312.

34. *Zion Golumbic E.M., Ding N., Bickel S. et al.* Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party", 2013. Neuron, 77 (5). Pp. 980-991.

REFERENCES

1. Abdel-Hamid O., Mohamed A., Jiang H., Penn G. Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition, Proc. IEEE Int. Conf. Acoust., Speech, Signal Process, 2012. Pp. 4277-4280.

2. Chomsky N.A. A Review of Skinner's Verbal Behavior, [Readings in the Psychology of Language]. Prentice-Hall, Upper Saddle River, New Jersey, 1967. Pp. 636.

3. Coates A., Ng A.Y. Learning feature representations with *k*-means. Neural Networks: Tricks of the Trade, 2012. Pp. 561-580.

4. De Mulder W., Bethard S., Moens M.-F. A Survey on the Application of Recurrent Neural Networks to Statistical Language Modeling // Computer Speech and Language. 2015. №30(1). Pp. 61-98.

5. De Saussure F. *Kurs obshei lingvistiki* [Course in General Linguistics]. *Yekaterinburg: Izdatel'stvo Ural'skogo Universiteta*, 1999. Pp. 256.

6. Deng L., Li X. Machine Learning Paradigms for Speech Recognition: An Overview, IEEE Transactions on Audio, Speech, and Language Processing, 2013. № 21(5). Pp. 1060-1089.

7. Gazzaniga M.S. Conversations in the Cognitive Neuroscience. The MIT Press, Cambridge, 1996. Pp. 752.

8. Ghai W., Singh N. Literature Review on Automatic Speech Recognition // International Journal of Computer Applications. 2012. № 41 (8). Pp. 42-50.

9. Ghitza O. Auditory nerve representation as a front-end for speech recognition in a noisy environment // Computer Speech and Language, 1986. Vol. 1. Pp. 109-130.

10. Gupta V. A Survey of Natural Language Processing Techniques // International Journal of Computer Science & Engineering Technology. 2014. № 5 (1). Pp. 14-16.

11. Haikonen P. The Cognitive Approach to Conscious Machines, imprint Academic, Exeter, UK, 2003. Pp.300.

12. Hinton G., Deng L., Yu D., et al Deep neural networks for acoustic modeling in speech recognition, IEEE Signal Process. Mag, 2012. № 29(6). Pp. 82-97.

13. Juan B.H. Speech Recognition in Adverse Environments // Computer Speech and Language, 1991. Vol. 5. Pp. 275-294.

14. Jurafsky D., Martin J. Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition, Prentice Hall, Boston, 2008. Pp. 1032.

15. Kotseruba I., Tsotsos J.K. A Review of 40 Years of Cognitive Architecture Research: Core Cognitive Abilities and Practical Applications, 2016 available at: arxiv.org/abs/1610.08602.

16. Mazurenko I.L. *Komp'utrenye sistemy raspoznavaniya rechi* [Computer Speech Recognition Systems] // *Intellektual'nye sistemy* [Intellectual Systems], 1998. Vol. 3. issue. 1-2. Pp. 117-134.

17. Minsky M. The Society of Mind. Simon and Shuster. New York, 1988. Pp. 336.

18. Mohamed A., Dahl G., Hinton G. Acoustic modeling using deep belief networks, IEEE Audio, Speech, Lang. Process, 2012. № 20(1). Pp. 14-22.

19. Morozov V.P., Vartanyan I.A., Galunov V.I. *Vospriyatiye Rechi: voprosy funktsional'noi asimmetrii mozga* [Speech Perception: Issues of functional brain asymmetry]. *Leningrad: Nauka,* 1988, 135 c.

20. Nagoev Z.V. *Intellektika, ili Myshlenie v zhivych i iskusstvennych sistemach* [Intellectics or Thinking in Living and Artificial Systems]. *Nalchik: Izdatel'stvo KBNC RAN,* 2013. Pp. 232.

21. Nagoev Z.V., Nagoeva O.V. *Izvlechenie znanii iz mnogomodal'nyh potokov nestrukturirovannyh dannyh na osnove samoorganizatsii mul'tiagentnoi kognitivnoi arhitektury mobil'nogo robota* [Knowledge Extraction from Multimodal Streams of Unstructured Data on the Base of Self-Organization of Multi-Agent Cognitive Architecture for Mobile Robot], *Izvestia KBNC RAN* [News of KBSC of RAS]. 2015. № 6 (68). Pp. 73-85.

22. Nagoev Z.V., Nagoeva O.V. *Zritel'nyi analizator intellektual'nogo robota dlya obrabotki nestrukturirovannyh dannyh na osnove mul'tiagentnoi neirocognitivnoi arhitektury* [Visual Analyzer of Intellectual Robot for Unstructured Data Processing on the Base of Multi-agent Neurocognitive Architechture] // *Perspektivnye sistemy i zadachi upravleniya: Materialy vserossiiskoi nauchno-prakticheskoi konferencii* [Advanced Systems and Management Tasks: Proceedings of the 12th All-Russia Conference]. Rostov-on-Don, 2017. Pp. 457-467.

23. Nagoev Z.V., Denisenko V.A., Lyutikova L.A. *Sistema obucheniya avtonomnogo sel'skohozyaistvennogo robota raspoznavaniyu staticheskih izobrazhenii na osnove multiagentnyh kognitivnyh arhitektur* [Learning System of Autonomous Agricultural Robot for Static Images Recognition on the Base of Multi-Agent Cognitive Architectures] // *Ustoichivoie razvitie gornyh territorii* [Sustainable Development of Mountain Territories]. 2018. № 2. Pp. 289-297.

24. Nagoev Z., Lyutikova L., Gurtueva I. Model for Automatic Speech Recognition Using Multi-Agent Recursive Cognitive Architecture, Annual International Conference on Biologically Inspired Cognitive Architectures BICA, 2018, Prague, Chech Republic http://doi.org/ 10.1016/j.procs.2018.11.089

25. Newell A. Unified Theories of Cognition, Harvard University Press, Cambridge, Massachusetts, 1990. Pp. 576.

26. Rabiner L.R., Schafer R.W. *Tsifrovaya Obrabotka Rechevyh Signalov* [Digital Processing of Speech Signals]. Moskva: Radio I Svyaz', 1981. Pp. 496.

27. Reddy R. Speech Recognition by Machine: A Review, Proceedings of the IEEE, 1976. № 64 (4). Pp. 501-531.

28. Ronzhin A.L., Karpov A.A., Kagirov I.A. *Osobennosti distantsionnoi zapisi i obrabotki rechi v avtomatah samoobsluzhivaniya* [Peculiarities of Remote Recording and Speech Processing in Self-Service Machines]. // *Informatsionno-upravlyayushie sistemy* [Information and Control Systems], 2009. № 5. Pp. 32-38.

29. Schunk D. H. Learning Theories: An Educational Perspective, Pearson Merrill Prentice Hall, Boston, 2011. Pp. 576.

30. Van Veen B.D., Buckley K.M. Beamforming: A Versatile Approach to Spatial Filtering, IEEE ASSP Magazine, 1988. № 5(2). Pp. 4-24.

31. Waibel A., Lee K.-F. Readings in Speech Recognition, Morgan Kaufman, Berlington, 1990. Pp. 680.

32. Wooldridge M. An Introduction to Multi-Agent Systems, Wiley, Hoboken, 2009. Pp. 366.

33. Zinder L.R. *Obshaya Fonetika* [General Phonetics]. *Moscow: Vysshaya Shkola*, 1979. Pp. 312.
34. Zion Golumbic E.M., Ding N., Bickel S., et al. Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party", 2013. Neuron, 77 (5). Pp. 980-991.

THE BASIC ELEMENTS FOR COGNITIVE MODEL OF SPEECH PERCEPTION MECHANISM ON THE BASE OF MULTI-AGENT RECURSIVE INTELLECT

Z.V. NAGOEV[1], I.A. GURTUEVA[2]

[1] Federal state budgetary scientific establishment "Federal scientific center
"Kabardin-Balkar Scientific Center of the Russian Academy of Sciences"
360002, KBR, Nalchik, 2, Balkarov street
E-mail: cgrkbncran@bk.ru
[2]Institute of Computer Science and Problems of Regional Management –
branch of Federal public budgetary scientific establishment "Federal scientific center
"Kabardin-Balkar Scientific Center of the Russian Academy of Sciences"
360000, KBR, Nalchik, 37-a, I. Armand St.
E-mail: iipru@rambler.ru

In this paper, the generalized architecture used in almost all modern systems of automatic speech recognition is analyzed. The necessity of developing a fundamentally new approach to solving speech recognition problems is outlined. A formal description of the structure of the speech perception act is proposed for use as a general theoretical basis in the development of universal automatic speech recognition systems that are highly effective in conditions of high noise and "cocktail party" situations. The general structural dynamics of the speech recognition process has been developed, which allows to take into account the linguistic and extra-linguistic aspects of a speech message. The concept of an articulation event as a minimal basic pattern of sound image recognition has been proposed. The recognition process is structured based on the functional determinants of the situation. The need to analyze the numerous sources of information accompanying the sound message, the rejection of the search for an invariant here is of fundamental nature. Multi-agent systems were chosen as the formal means for implementation. Multi-agent approach allows to differentiate and analyze sounds of different nature. This makes the proposed model unique and gives it advantages in the so-called "cocktail party" situation, as well as in tasks where the noise level is extremely high.

Keywords: artificial intellect, multi-agent systems, speech recognition, artificial neural networks.

Работа поступила 05.06.2019 г.