В этом легко убедиться. Такую же постоянную величину можно получить для подобных треугольников, круга и, скорее всего, для большого множества характерных по своему качественному содержанию фигур. Но если предположить подобную манипуляцию со звуковым потоком, то возникнут те же вопросы: где начало такой условной фигуры, где ее завершение.

Новая качественная мера сопоставления волн

Разумно обратиться к человеческому распознаванию в противовес техническому [3–5]. Рассуждение лучше вести последовательно и внимательно, пытаясь глубоко разобраться в нюансах человеческого восприятия окружающего мира и распознавания звуков, предметов и т.д. За что цепляется человеческое сознание при распознавании окружающего мира?

- Вряд ли человек обозревает мир линейно и последовательно. Противоположность такому обозрению – ***объемность*** *и* ***параллельность***. Линейность возможна тоже, но на фоне объемного восприятия и последовательного движения на фоне параллельного поступления информации в случае фокусирования внимания на определенном предмете или звуке.
- Вряд ли возможно запечатлеть окружающий мир, если записывать его содержание непрерывно. Значит, мир воспринимается дискретно и очень экономно при бессознательном ощущении ограниченности ресурсов.
- Воспринимаемый мир человек запечатлевает иерархически и только в рамках существенных деталей (опять экономия), поднимая на более высокий уровень памяти и сознания наиболее значимые, переломные моменты.
- Образ о предмете, с которым сравниваются прообразы предметов, хранится в ***сжатой форме***, не зависящей от числового значения параметров, на вершине иерархии знаний о предмете и в окружении существующего объема **связей** с другими образами.

Не исключено, что есть и другие характерные моменты человеческого распознавания. Но этого пока достаточно, чтобы, исходя из особенностей такого распознавания, перейти к техническому распознаванию.

Во-первых, параллельность и объемность требуют нахождения характерного «кадра» и способа его выделения. Следует отметить, что подобные кадры должны иметь свойство воспроизводства (тиражирования) в других звуковых потоках, порожденных другими личностями, и в то же время повторяться в конкретном звуковом потоке, чтобы обеспечить точность восприятия.

Во-вторых, дискретности, вытекающей из самого процесса оцифровки звукового потока, недостаточно. Часть информации является несущественной, и достаточно наличия характерных переломных точек, по которым можно восстановить промежуточную информацию. Кроме этого, они могут стать реперными точками установления длины участков (в количестве отсчетов) повторения подобной характерной последовательности точек. Таким образом, получается большая экономия при обработке звукового потока без потери качества содержания.

В-третьих. Учитывая вложенность волн и не учитывая (схлопывая) более слабые волны, можно получить следующий уровень иерархии отношений, где уменьшается объем характерных точек. Повторяя такую процедуру, получаем процесс, сходящийся к одной из фигур, указанных в работе [6].

В-четвертых, чтобы получить характерную картину на качественном уровне сравниваемых участков волны, которые не зависели бы, в первую очередь, от длины волны, следует обратиться к отношениям уже множества характерных точек. В работе [7] отмечена значимость в установлении отношений упорядоченных последовательностей, которая позволяет поделить последовательности оцифрованных значений чисел на качественные классы множеств.

Все возможные последовательности, состоящие из трех любых значений чисел, можно разбить на 13 классов подмножеств с учетом комбинаций их возможных отношений. И такой набор базовых отношений назовем ***примитивами***. Подобная ограниченность количества классов подмножеств наблюдается при дальнейшем увеличении точек рассмотрения и весьма ощутимо уменьшает количество комбинаций. Фактически такие отношения отношений определяют конфигурацию, или фигуру отношений. Интуитивно понятно, что сравнивать имеет смысл близкие по структуре фигуры отношений, что синхронизирует (совмещает) изображения при одинаковом количестве выбранных характерных точек. Но и число таких рассматриваемых точек должно быть в разумных пределах, чтобы обеспечить и обнаружить повторяемость конфигурации, так как здесь при фиксированном потоке действует закон: чем больше точек, тем меньше вероятность заметить подобные повторяющиеся по конфигурации участки.

В теоретических рассуждениях все вроде бы понятно и логично, но как сделать шаг в сторону практического применения и «запечатления» конфигураций? Была предложена широко используемая форма представления в виде матрицы, которую авторы назвали ***структурной матрицей***. Совокупность и иерархия таких структур определяют общую конфигурацию звукового потока. Она оказалась той качественной мерой**,** которая позволяет разбивать звуковой поток на иерархию структур, сравнивать такие структуры, а также, привлекая

числовую меру, дополнять и углублять далее процесс распознавания. В подобной раскладке отсутствует проблема совмещения характерных участков, так как построение матриц производится с учетом всей иерархии построения потока и наряду с характерными точками числовой меры учитываются точки с качественной мерой, фактически отражающей спектральную характеристику, которая содержится в характерной точке. В структурной матрице картина отношений будет характеризоваться некоторыми последовательностями характеристик $\chi_{i,j}$, которые фактически характеризуют конфигурацию отношений. Без потери общности возьмем общеизвестное множество отношений, состоящее из трех характеристик и пустого значения ⊗: $\chi \in \{\otimes, <, >, =\}$.

Обозначим $a_{i,j}$ элемент матрицы, который может принимать одну из символьных характеристик отношений элементов i,j. Матрицу будем представлять в виде числового идентификатора от 0 до 3 соответственно. В силу упорядоченности и отсутствия необходимости рассмотрения отношений объектов в обратном порядке будем пользоваться только верхней бездиагональной треугольной матрицей. Бездиагональная *верхнетреугольная матрица* (ВТМ) – это квадратная матрица, у которой всегда все элементы ниже главной диагонали и по диагонали (0 диагональ) являются пустыми ($a_{ij} = \otimes$ при $i \geq j$). Все остальные элементы могут иметь любое значение из указанного выше списка. При этом нулевое значение выше диагонали будет означать, что пока это отношение не выведено либо не установлено. Процесс установления отношений проходит в три этапа:

– структуризация – получение отношений в виде примитивов;

– выведение некоторых отношений на основе полученных примитивов по правилам выведения отношений;

– явное установление отношений для отношений, не выводимых неявно, предполагающее процедуру явного сравнения двух количественных характеристик в структурированном потоке с последующим повторением этапа выведения.

Если в качестве идентификатора объекта отношений использовать их номера, а отношения между ними фиксировать как некоторое значение на пересечении строк и столбцов, то можно представить всю совокупность отношений в виде ВТМ (рис. 1).

В данной матрице отношений имеются два вида нумерации: как элемент стандартной квадратной матрицы $a_{i,j}$ – матричная индексация и последовательная внутренняя индексация, указанная в индексе вопроса «**?**$_i$». Две диагонали вверх заполняются на основе примитивов, получаемых в процессе первичной структуризации. Таким образом, первая и вторая диагонали будут заполнены изначально.

i/j	0	1	2	3	4	5	6	…	n-1	n
0	⊗	**a**$_{01}$	**a**$_{02}$	**?**$_1$	**?**$_3$	**?**$_6$	**?**$_{10}$	…	•	•
1		⊗	**a**$_{12}$	**a**$_{12}$	**?**$_2$	**?**$_5$	**?**$_9$	…	•	•
2			⊗	**a**$_{23}$	**a**$_{24}$	**?**$_4$	**?**$_8$	…	•	•
3				⊗	**a**$_{34}$	**a**$_{35}$	**?**$_7$	…	•	•
4					⊗	**a**$_{45}$	**a**$_{46}$	…	•	•
5						⊗	**a**$_{56}$	…	•	•
6							⊗	…	•	•
…								⊗	•	•
n-1									⊗	**a**$_{n-1,n}$
n										⊗

Рис. 1. Матрица отношений между упорядоченными объектами

Fig. 1. Matrix of relations between ordered objects

Необычный вариант внутренней индексации снизу вверх, а затем вправо и опять снизу вверх объясняется последовательностью продвижения по неизвестным значениям отношений, которое удобно использовать для выведения некоторых первоначально неизвестных значений элементов матрицы, исходя из предыдущих известных значений отношений.

Полностью заполненная матрица отражает качественную характеристику упорядоченной последовательности значений объектов или конфигурацию последовательности объектов безотносительно их численного значения. Можно сказать, что это матрица полной связанности объектов (каждый с каждым) с установленным характером таких отношений в виде принятых условных значений. Это весьма важная характеристика для определения подобия объектов без привязки к их размерности, что и соответствует распознаванию объектов. Содержательно первая диагональ от центральной диагонали устанавливает отношения между соседними элементами, вторая – отношения между элементами, идущими через раз, третья – через два и т.д. Строка матрицы отражает отношения со всеми далее идущими элементами характерных точек.

Фундаментальное значение такого представления состоит в том, что в практическом плане появляется качественная мера, которая дает возможность качественного сравнения различных объектов через сравнение таких матриц. Более того, при такой связанности объектов отношениями число возможных представлений матриц и, соответственно, отношений, оказывается, весьма ограничено по отношению к возможным комбинациям заполнения матрицы. Содержательно все это отражает факт ограниченности качественных представлений связанных событий и безграничность их численного представления. Именно подобная огра-

ниченность и распознавание через подобие защищают наше мышление от бескрайности фактического потока информации из окружающего мира.

Качественное рассмотрение задачи распознавания фонем

Задачу распознавания речи будем рассматривать не как нечто только формальное и каноническое, а с позиций того, что, в принципе, существуют два вида такого распознавания: человеческое и техническое. Вникая в сущность человеческого распознавания, можно черпать из нее идеи для технического распознавания, учитывая это в применяемых алгоритмах. Техническое распознавание – это распознавание сущности физических значений величин техническими средствами и алгоритмами, которое удовлетворяет определенному качеству распознавания и устанавливается человеком как некий неформальный набор требований. Следует подчеркнуть, что сугубо формальная математическая постановка задачи, присутствующая в технической литературе и статьях, подобна идее вечного двигателя: найти метод и формулу, которая позволит определять объект с точностью до очень малого ε относительно распознаваемого объекта.

Однако следует особо отметить, что переход к задачам человеческого распознавания ведет в область качественных задач [5, 6], где отсутствует нечто точное и даже нечеткое, а есть то, что имеет допустимый качественный интервал или качественную характеристику, которая приводит к положительному результату и соответствует требованиям практики. Распознавание относительно человека – это итерационный и многоканальный процесс на качественном уровне, который прерывается по мере достаточности информации для принятия решения. И это не каприз в нежелании углубляться дальше, а закон сохранения энергии и памяти, который почему-то иногда отождествляется со словом «лень». С другой стороны, для распознавания чего-то необходимо, чтобы некоторая первооснова повторилась несколько раз и была устойчивой в виде образа некоторый промежуток времени. Рассмотрим звуковой поток амплитуд в виде осциллограммы и попробуем подтвердить приведенные рассуждения.

На рисунке 2, где по оси X указаны отсчеты, а по оси Y – значения амплитуд, представлена осциллограмма участка речи для фонемы «а», в которой характерные точки пронумерованы и выделены. Между ярко выраженными характерными переломными точками (например, 40 и 140 – отчет для минимума) наблюдаются одинаковые очертания рисунков, которые являются *исходными фонемными первоосновами* и должны повторяться не менее 10–15 раз (проверено экспериментально) для того, чтобы можно было уловить содержание соответствующей фонемы. Первое, что следует заметить, почти секундное звучание отдельно взятого слова средней длины выливается при записи в десятки тысяч значений амплитуд, а в сознании человека это мгновенный образ и всего несколько письменных знаков текста.

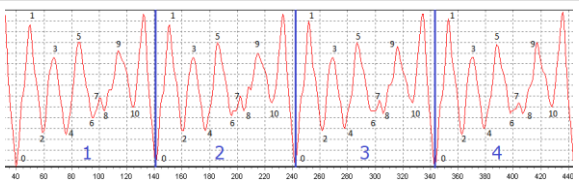

Рис. 2. Голосовые фонемные кванты фонемы «а»

Fig. 2. Voice phonemic quanta of "a" phoneme

С технической точки зрения для получения звучания фонемы требуется получить формальное представление подобной первоосновы и повторить ее необходимое количество раз. Нечто похожее будет относиться и к другим фонемам. Тогда из множества подобных формальных представлений техническим знаком – буквой фонемы будет номер варианта фонемного примитива (i – номер или вариант рисунка, соответствующий фонеме «а», повторенный N раз). На этом уровне поток сигналов может быть свернут в знаковую первооснову, по которой можно повторить техническое воспроизведение фонемы (обратная операция по отношению к распознаванию). Для представления первоосновы была предложена модель представления таких участков, как это было сказано выше, в виде матрицы отношений, предопределяющей содержание фонемной первоосновы без привязки к значениям амплитуд. Добавление числовой информации: значение количества отсчетов между характерными точками и самих значений амплитуд превращает матрицу в функциональную, по которой легко восстановить все значения амплитуд для каждой точки отсчета. Звучание фонемы можно получить и без наличия значения амплитуд (их можно сформировать по некоторому алгоритму, сохраняя конфигурацию и количество отсчетов), и представление «матрица плюс значения количества отсчетов между характерными точками» будем называть потенциальной матрицей.

Звуковые колебания в промежутках между ярко выраженными характерными точками являются фонемными квантами, совокупность таких участков, в которых проявляется фонема, будем называть фонемным примитивом. Сам промежуток (длина волны голоса) без фонемного кванта фактически является первоосновой голоса, произносящего фонему. Этот промежуток будем называть голосовым квантом, а совокупность голосового кванта с фонемным квантом – голосовым фонемным квантом. Таким образом, с необходимостью следует понятие голосового фонемного примитива. И вся наша речь состоит, как из кирпичиков, из та-

ких голосовых фонемных примитивов, всегда окрашенных голосом говорящего.

Рассмотрим отношения между характерными точками без акцентирования внимания на величине таких отношений и введем одинаково интерпретируемые обозначения: ⊗ (пусто), < (меньше), > (больше) и = (равно) и обозначения их цифрами 0, 1, 2 и 3 соответственно. Каждая цифра, таким образом, отражает вполне конкретное содержание, и для ее фиксации будет достаточно двухбитового поля.

На этой основе продемонстрируем подобие фонемных квантов. Используя введенные соглашения, составим формальную картину, описывающую все отношения характерных точек между собой на фонемном кванте. И затем эту процедуру повторим для каждого фонемного кванта (рис. 1).

Составим матрицу отношений для каждого фонемного кванта между характерными точками, используя приведенные выше обозначения. Берем точку 0, сопоставляем ее с другими точками (1, 2, 3 и т.д.) и установленные отношения в виде цифр записываем в нулевую строку таблицы. Точно так же поступаем с точкой, обозначенной цифрой 1, отношения которой будут записаны во вторую строчку. При этом не требуется установление обратных отношений с точкой 0, так как они противоположны по отношению точки 0 к точке 1, кроме случая равенства. Такую процедуру проделаем для всех 10 точек. На рисунке 3 показана структурная матрица отношений 1-го фонемного кванта. Ту же последовательность действий проведем для 2-го фонемного кванта и получим матрицу отношений для фонемного кванта (рис. 4).

Построенные матрицы практически одинаковы относительно их отношений, кроме одного, выделенного на рисунке в соответствующей точке. Проведем такую же работу для 3-го и 4-го фонемных квантов. Здесь появляются две новые дополнительные несовпадающие клетки (на рисунках 5 и 6 выделены тоном).

Не останавливаясь на причинах таких изменений в выделенных точках, можно утверждать, что матрицы практически одинаковы.

Далее сопоставим количество отсчетов между характерными точками (см. таблицу).

Сопоставление отсчетов между характерными точками звукового потока

Comparison of samples between audio stream characteristic points

Номер кванта	Номер точки											Сумма
	0	1	2	3	4	5	6	7	8	9	10	
1		9	8	9	9	10	5	4	3	15	8	80
2		9	8	9	9	10	10	6	3	16	8	88
3		10	7	9	9	10	3	12	10	11	7	88
4		10	7	9	9	10	10	6	10	11	8	90
Среднее:		9	8	9	9	10	7	7	7	13	8	85-86

Данное исследование было проведено на структурированном потоке фонем «а» и показало, что с точностью менее 10 % наблюдается расхождение в структурных матрицах. Для отсчетов расхождение на единицу можно считать незначительным. Исходя из среднего значения изменения интервала между характерными точками, оно колеблется у отметки в 8 отсчетов. Любой сдвиг на несколько отсчетов влево или вправо любого из участков уже не позволит получить такую картину.

Таким образом, продемонстрирована потенциальная возможность сравнения подобных участков осциллограммы на основе матриц отношений между собой – это первая качественная характеристика при рассмотрении потока. Количество отсче-

1	1	1	1	1	1	1	1	1	1	
⊗	2	2	2	2	2	2	2	2	2	
	⊗	1	2	1	1	1	1	1	1	
		⊗	2	1	2	2	2	1	2	
			⊗	1	1	1	1	1	1	
				⊗	2	2	2	2	2	
					⊗	1	1	1	1	
						⊗	2	1	2	
							⊗	1	1	
								⊗	2	
									⊗	

Рис. 3. Матрица отношений для 1-го голосового фонемного кванта

Fig. 3. A relational matrix for the 1st voice phonemic quantum

1	1	1	1	1	1	1	1	1	1	
⊗	2	2	2	2	2	2	2	2	2	
	⊗	1	3	1	1	1	1	1	1	
		⊗	2	1	2	2	2	1	2	
			⊗	1	1	1	1	1	1	
				⊗	2	2	2	2	2	
					⊗	1	1	1	1	
						⊗	2	1	2	
							⊗	1	1	
								⊗	2	
									⊗	

Рис. 4. Матрица отношений для 2-го голосового фонемного кванта

Fig. 4. A relational matrix for the 2nd voice phonemic quantum

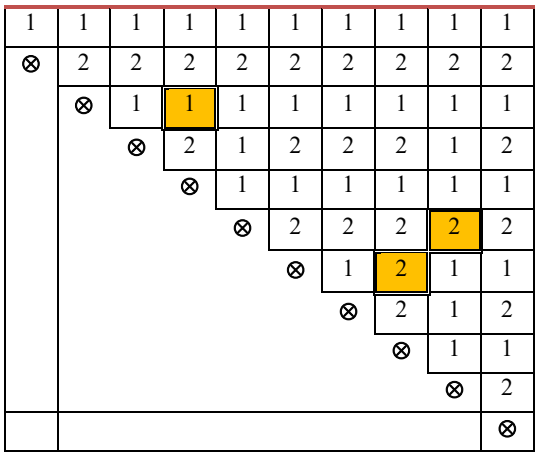

Рис. 5. Матрица отношений для 3-го голосового фонемного кванта

Fig. 5. A relational matrix for the 3rd voice phonemic quantum

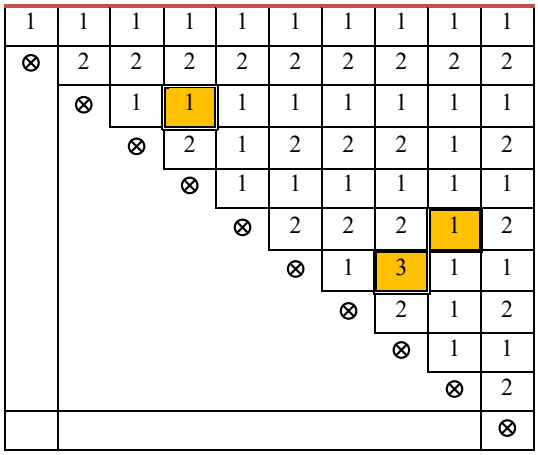

Рис. 6. Матрица отношений для 4-го голосового фонемного кванта

Fig. 6. A relational matrix for the 4th voice phonemic quantum

тов между характерными точками – второй набор параметров, который определяет фонемный квант. И третий, необязательный набор – это значения амплитуд для характерных точек. В матрицах отношений из 55 значений в каждой матрице различаются только три позиции (5,4 %).

Введение структурных матриц потребовало кардинального пересмотра инструмента исследований, который должен был учесть новые подходы в извлечении информационного содержания волн.

Создание системы визуализации и редактирования цифрового потока сигналов на основе их структурного представления

Для исследования сигналов уже не один десяток лет применяется визуализация поступающего потока сигналов в виде осциллограммы [8–10]. Последнее время в качестве исходного материала для визуализации стали применять не сам сигнал, а его цифровое значение. Это привело к появлению так называемых программных осциллографов, и коррекция сигнала вылилась в изменение его значения. Сама осциллограмма – это источник анализа информационного содержания сигнала через визуальное восприятие человеком. Объемная картина всего потока сигналов, масштабирование различных участков, повторение и другие возможности оказывают существенную помощь в анализе сигналов человеком. Но наступил новый этап в расширении возможностей по извлечению информационного содержания сигнала – распознавание, ориентированное на те или иные предметные области. Особенно популярна область исследований по распознаванию речи. Независимо от предметной области исследований, включая медицину, сейсмику, связь и другие, любой поток сигналов можно рассматривать в рамках указанного подхода, так как в их основе лежит волновой характер, с одной стороны, повторяемость отдельных значений, а на качественном уровне характерная форма участков таких участков – с другой. Таким образом, и на содержание поступающего потока сигналов теперь уже в цифре можно посмотреть под разным углом зрения.

Безусловно, примитивный взгляд на информационное содержание сигналов как не связанное какой-то структурной моделью позволяет только визуализировать содержание, которое доступно обработке человеком. Заманчивой перспективой является возможность нащупать такое содержание на уровне некоторых структурных моделей и, пользуясь алфавитом таких характерных структур, вести исследование и анализ любых сигналов. Именно такой подход, основанный на структуризации цифрового потока сигналов и заявленный как логико-лингвистический подход, продвигается в исследованиях по распознаванию речи [6].

В анализе речевых потоков в этих исследованиях интенсивно использовался фактически цифровой осциллограф, разработанный Малковым М.А., а вскрываемые структуры хранились в виде отдельных файлов и затем переводились для визуализации в последовательный поток сигналов. В качестве следующего шага требуется перейти на новый уровень так называемого алфавита структур и производить редакцию не отдельных сигналов или отдельных их групп, а их алфавитной последовательности. Таким оообразом, по отношению к предыдущему инструменту, а также при наличии широкого ассортимента программных осциллографов и звуковых редакторов ключевыми отличительными особенностями нового редактора стали следующие:

– ориентация в исследовании алгоритмов на структуры звукового потока с целью эффективного и качественного распознавания речи;

– использование логико-лингвистического подхода, позволяющего структурно представить звуковой поток;

– одновременная ориентация исследований как на распознавание, так и на синтез при обработке речи;

– возможность использования нового подхода, алгоритмов и структур для других видов колебаний;

– обеспечение исследования доказательствами применимости предлагаемой структуризации для распознавания и синтеза речи в виде фиксируемых сеансов;

– создание условий комбинирования применяемых методов обработки через API-функции и с исходным «алфавитом» структур в виде сценария включаемых функций и структур;

– включение средств редактирования (групповое амплитудное изменение, групповое изменение количества отсчетов между характерными точками, групповая замена одного структурного символа на другой) в рамках новых структурных единиц и отработка в таком представлении основных базовых единиц (фонемных квантов), образующих фонемы в результате их многократного повторения;

– формирование цепочек фонемных квантов и цепочек фонемных примитивов для формирования отдельно взятого слова или фразы.

При рассмотрении звукового потока будем придерживаться следующей концепции структуризации:

– информационное содержание волны – это не точная математическая функция, а набор некоторых характерных элементарных фрагментов (очертаний);

– содержание слова, извлеченного из волны, – это структура структур таких элементарных фрагментов;

– элементарный фрагмент не зависит от значения амплитуды;

– определенный разброс числовых характеристик как для расстояний между характерными точками, так и для значений амплитуд.

Подтверждение данного подхода будет происходить за счет структуризации звукового потока на основе примитивов, позволяющих выделить характерные точки потока и структурных матриц (по факту это сжатие с потерями), и деструктуризации звукового потока, представленного на основе примитивов и структурных матриц, а затем воспроизведении восстановленного потока для оценки качества восстановленного сигнала.

Редактор должен обеспечивать следующий набор функций:

– запись звука;

– воспроизведение звука;

– фиксация всей записи звуков или выделение отдельных участков звукового потока и их запись в виде файла;

– считывание уже записанного звукового файла;

– считывание (запись) звуковых единиц в виде функциональных и структурных матриц;

– считывание (запись) восстановленных звуковых файлов;

– считывание (запись) измененных, скорректированных звуковых файлов;

– графическая визуализация звукового потока в виде визуального представления с возможностью визуализации нескольких потоков одновременно;

– воспроизведение звука по выделенному участку визуального представления;

– масштабирование визуального представления;

– изменение, коррекция (редактирование) функциональных и структурных матриц и примитивов.

Схема взаимодействия компонент действующей системы показана на рисунке 7. В ее составе следующие элементы:

– API-функции алгоритмов структуризации и деструктуризация звукового потока данных, работы с матрицами и т.п., которые находятся в специальной библиотеке API STRUCT;

– БД структурных матриц;

– библиотека потенциальных матриц – фонемных квантов;

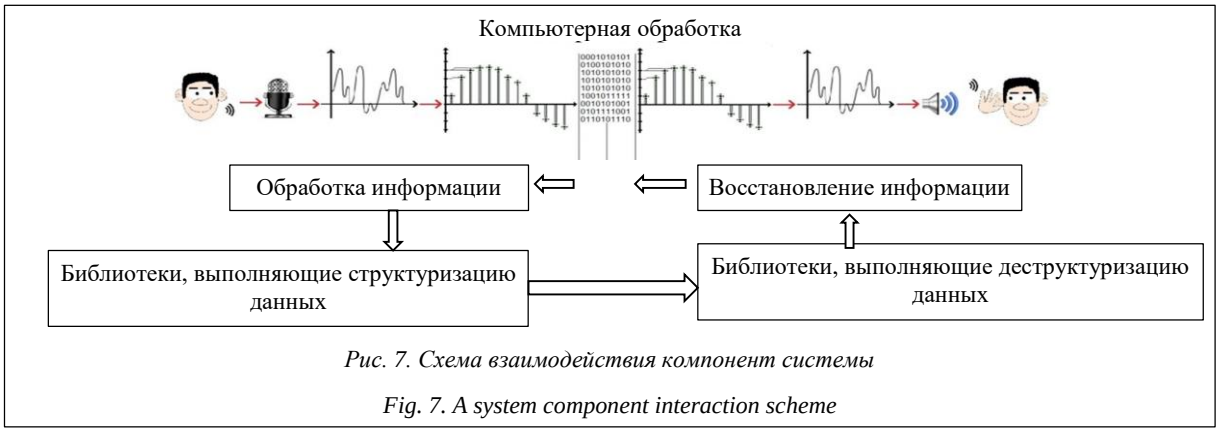

Рис. 7. Схема взаимодействия компонент системы

Fig. 7. A system component interaction scheme

– библиотека сценариев включаемых компонент;

– подсистема проверки интерфейсов.

Созданный для взаимодействия с ранее упомянутыми библиотеками и БД *структурных матриц отношений* интерфейс позволяет автоматизировать процессы структуризации и деструктуризации и отрабатывать цепочки исследований по замкнутой схеме: анализ–синтез речи.

Редактор может сформировать исходный для исследования материал, используя микрофон, записи звука из других источников через звуковые файлы, через выделение интересующих фрагментов и оформление его отдельной записью. В процессе исследования формируется множество промежуточных структур, которые также могут быть исходным материалом. Таким образом, исходные данные для редактирования в системе могут поступать из следующих источников: запись речи через микрофон; звуковой файл типа WAV, записанный через другие устройства; восстановленный звуковой поток; упакованный звуковой поток; БД структур звуковых единиц.

Выходной информацией редактора могут быть звуковой поток в структуре типа WAV и в виде примитивов и структурных матриц.

Наряду с этим фиксируется служебная информация: шаблоны сценариев, исполняемые сценарии, дублирующие ярлыки звуковые файлы и т.д.

Базовое ядро, то есть основа системы, представляет собой:

– исходный код программы со всеми интерфейсами взаимодействия с человеком, написанный на основе открытого кода инструментов – языке C;

– библиотеку функций в виде DLL-библиотеки на ассемблере FASM, разработанную Н.Е. Балакиревым;

– окно приложения и все его графические элементы, реализованные при помощи графической библиотеки GLEW и библиотеки GLFW;

– базовые структуры, разрабатываемые для проекта распознавания речи.

Интерфейс взаимодействия включает обращение через клавиатуру, мышку (предполагается и речевое взаимодействие), а ответная часть – множество преобразованных экранных форм и звуковых файлов с поддержкой голосового сопровождения нажимаемых клавиш и кнопок. Работа может производиться одновременно с четырьмя звуковыми потоками во время одного сеанса исследований.

Литература

1. Петровский А., Борович А., Парфенюк М. Обработка речи на основе дискретного преобразования Фурье с неравномерным частотным разрешением // Речевые технологии. 2008. № 3. С. 3–15.

2. Музычук Д.С., Медведев М.С. Сегментация, шумоподавление и фонетический анализ в задаче распознавания речи // Молодой ученый. 2013. № 6. С. 86–96.

3. Гренандер У. Лекции по теории образов: Анализ образов; [пер. с англ.]. М.: Мир, 1981. 448 с.

4. Гренандер У. Лекции по теории образов: Регулярные структуры; [пер. с англ.]. М.: Мир, 1983. 432 с.

5. Grenander U. A calculus of ideas: a mathematical study of human thought. World Scientific Publi., 2012.

6. Балакирев Н.Е. Логико-лингвистический подход при обработке колебательных сигналов (базовая концепция) // Информатика: проблемы, методология, технологии: матер. XIV Междунар. конф. Воронеж: 2014. Т. 1. С. 331–335.

7. Балакирев Н.Е. Количественная и качественная оценка исследуемых объектов на примере простейших отношений // Вестн. ВГУ. Сер.: Системный анализ и информационные технологии. 2016. № 2. С. 65–72.

8. Карпов А.А. Реализация автоматической системы многомодального распознавания речи по аудио- и видеоинформации // Автоматика и телемеханика. 2014. Т. 75. № 12. С. 125–138.

9. Schwarz P. Phoneme recognition based on long temporal context. Ph.D. Thesis, Brno Univ. of Technology, 2008.

10. Bailly G., Perrier P., Vatikiotis-Bateson E., Audiovisual speech processing. Cambridge Univ. Press, 2012, 506 p.

Software & Systems
DOI: 10.15827/0236-235X.124.768-776

Received 14.09.18
2018, vol. 31, no. 4, pp. 768–776

Structuring and qualitative consideration of a audio stream in a speech synthesis and analysis system

N.E. Balakirev [1], *Ph.D. (Engineering), Professor, balakirev1949@yandex.ru*
H.D. Nguyen [1], *Postgraduate Student, nguyenhoangzuy@gmail.com*
M.A. Malkov [1], *Ph.D. (Engineering), Engineer-Programmer, maksi-malkov@yandex.ru*
M.M. Fadeev [1], *Student, _mix@bk.ru*

[1] *Moscow Aviation Institute (National Research University), Moscow, A-80, GSP-3, 125993, Russian Federation*

Abstract. Many approaches to speech recognition do not exclude new views on the recognition process and implementation. The paper focuses on presenting the idea of a new approach, although many practical results have already been obtained on its basis. It is practical results that led to a revision of the basic concepts of structuring a signal sound flow and further rethinking of the toolkit, which is based on a traditional oscillograph.

The most important part of a new model is a structural matrix. Such introduction is justified on the example of processing one phoneme. Structural matrices make it possible to obtain the information content of a sound stream at different levels and

with different recognition purposes: recognizing a phoneme, the voice of a particular person and speech tonality, including the tonality of the Southeast Asia phonemes.

As for the toolkit, an important step in visual representation, which remained unchanged externally, was the possibility of unlimited scaling and using GPU for such implementation. However, the main innovation is that the source material are the structures representing a special kind of model representation and demonstrating the asserted idea and the notion of a qualitative representation of an input signal flow. Systematization of the flow through the structure optimized the process of opening the content of constituent elements (primarily speech). However, it complicated processing to obtain the required material.

The toolkit involves both sides of work on the content of speech: analysis and synthesis. Hence, there are two processing directions: formation of structure models from a digitized signal and inverse transformation of such structures into an audio stream.

Keywords: recognition, technical speech recognition, qualitative measure, structural matrix, functional matrix, potential matrix, phonemic quantum.

References

1. Petrovsky A., Borovich A., Parfenyuk M. Speech processing based on discrete Fourier transformation with non-uniform frequency resolution. *Speech Technologies*. 2008, no. 3, pp. 3–15 (in Russ.).

2. Muzychuk D.S., Medvedev M.S. Segmentation, noise reduction and phonetic analysis in the problem of speech recognition. *Young Scientist*. 2013, no. 6, pp. 86–96 (in Russ.).

3. Grenander U. *Lectures in Pattern Theory. Vol. 2: Pattern Analysis*. Springer-Verlag Publ., NY, 1978 (Russ. ed.: Moscow, Mir Publ., 1981, 448 p.).

4. Grenander U. *Lectures in Pattern Theory. Vol. 3: Regular Structures*. Springer-Verlag Publ., NY, 1981 (Russ. ed.: Moscow, Mir Publ., 1983, 432 p.).

5. Grenander U. A *Calculus of Ideas: a Mathematical Study of Human Thought*. World Sci. Publ., 2012.

6. Balakirev N.E. A logical and linguistic approach in oscillatory signal processing (basic concept). *Computer Science: Problems, Methodology, Technologies: Proc. 14th Intern. Conf*. Voronezh, 2014, vol. 1, pp. 331–335 (in Russ.).

7. Balakirev N.E. Quantitative and qualitative assessment of the studied objects by the example of the simplest relations. *Proc. of Voronezh State Unive. Ser. Systems Analysis and Information Technologies*. 2016, no. 2, pp. 65–72 (in Russ.).

8. Karpov A.A. An automatic multimodal speech recognition system with audio and video information. *Autom. Remote Control*. 2014, vol. 75, no. 12, pp. 125–138 (in Russ.).

9. Schwarz P. Phoneme *Recognition Based on Long Temporal Context*. Ph.D. Thesis, Brno Univ. of Technology Publ., 2008.

10. Bailly G., Perrier P., Vatikiotis-Bateson E. *Audiovisual Speech Processing*. Cambridge Univ. Press, 2012, 506 p.

Примеры библиографического описания статьи

1. Балакирев Н.Е., Нгуен Х.З., Малков М.А., Фадеев М.М. Структуризация и качественное рассмотрение звукового потока в системе синтеза и анализа речи // Программные продукты и системы. 2018. Т. 31. № 4. С. 768–776. DOI: 10.15827/0236-235X.124.768-776.

2. Balakirev N.E., Nguyen H.D., Malkov M.A., Fadeev M.M. Structuring and qualitative consideration of a audio stream in a speech synthesis and analysis system. *Software & Systems*. 2018, vol. 31, no. 4, pp. 768–776 (in Russ.). DOI: 10.15827/0236-235X.124.768-776.