

# Технология распознавания и озвучивания текстов для обеспечения учебного процесса людей с нарушениями зрения

*Юрьев Г.А., аспирант Московского городского психолого-педагогического университета*



*Куравский Л.С., доктор технических наук, профессор*



В статье рассматривается способ решения проблемы доступа к текстовой информации, возникающей у людей с нарушениями зрения. Текст — основной источник знаний в пространстве, которое принято называть «информационным». Несмотря на огромное количество электронных источников, в настоящее время существует распространённая ситуация, когда доступен только бумажный вариант текста. Далеко не всегда нужную информацию можно получить из книг, набранных Брайлевским шрифтом. В то же время людям с нарушениями зрения остаются недоступными обычные печатные материалы и часть информации на экранах компьютерных мониторов. Поэтому разработка программно-аппаратных средств, обеспечивающих для незрячих людей доступ к этим источникам, является в настоящее время чрезвычайно актуальной.

В этой статье представлена новая технология обработки текстов для людей с нарушениями зрения, интегрирующая средства сканирования, распознавания и озвучивания.

• распознавание образов • нейронные сети • вейвлет-преобразования • сети Хэмминга • восстановление изображений • цепи Маркова.

The technology intended for recognition of text information by visually impaired persons is under consideration. Participation of such specialists in diagnostic works is frequently necessary because of social reasons and entails relevant technical support. The presented technology is intended for transformation of initial text images obtained online with the



aid of a video camera from computer monitors and similar sources to their audio representation. Initial images after their acquisition are converted into corresponding contour shapes by a proper wavelet transform and, then, decomposed into character strings, which are recognized by means of a set of techniques, one of which combines capabilities of wavelet transforms and relaxation neural networks. Recognized character strings are then spoken by a speech synthesizer. Features of the implemented hardware-software system are given.

- *pattern recognition* • *neural networks* • *wavelet transform* • *Hamming network*
- *restoring images*.

## Введение

Лёгкий доступ к информации стал неотъемлемой частью жизни. Для большинства людей получение информации не представляет труда. Однако слабовидящие люди лишены возможности пользоваться информацией в полном объёме. Проблема доступа к текстовой информации — одна из наиболее значимых в процессе адаптации [1] людей с нарушениями зрения в современной компьютеризированной среде. В настоящее время большое количество художественной литературы появляется в аудиоварианте. Однако ни публицистика, ни большая часть технической литературы в аудиоформате не выходят. У издательств в этом нет особой необходимости, так как аудитория подобных проектов была бы слишком мала. Но возможность для слабовидящих, например, в индивидуальном порядке прослушать газету, журнал, техническую инструкцию может быть весьма ценной.

В этой статье представлена новая технология обработки текстов для людей с нарушениями зрения, интегрирующая средства сканирования, распознавания и озвучивания. Она может быть использована для чтения литературы, изданной традиционным печатным способом, и работы за экраном компьютера (здесь и далее подразумевается любой жидкокристаллический дисплей) с обычным программным обеспечением, не предназначенным для незрячих пользователей.

Решение проблемы чтения именно плоскочечного текста обеспечивает доступ к образовательным ресурсам, повышает эффективность процесса обучения и позволяет реализовать профессиональные навыки. Распространённые традиционные средства доступа, включая азбуку Брайля, мультимедийные книги и программное обеспечение для озвучивания представленных в электронной форме текстов, частично решают указанную проблему. Эти средства требуют значительных усилий, затрачиваемых на предварительную подготовку исходного материала и перевода его в доступную для восприятия брайлевскую или электронную форму представления. Технология обеспечивает возможность считывания текста с бумажных и других носителей при минимальных требованиях к аппаратной поддержке, что повышает мобильность и доступность системы. Разработанное программное обеспечение включает развитые средства устранения ошибок, обусловленных низким качеством обрабатываемого изображения.

Из средств воспроизведения текстов, предназначенных для людей с нарушениями зрения, в настоящее время доступны:

- системы оптического распознавания текстов FineReader [6], CuneiForm [8] и другие, позволяющие озвучивать занесённые в память компьютера отсканированные тексты с бумажных источников;
- системы незрительного доступа JAWS [7], Adriane Knoppix [9] и их аналоги, озвучивающие представленные в электронной форме фрагменты текстов и имеющие развитую систему навигации по ключевым словам;
- звуковые книги, подготовленные в специальных аудиоформатах.

Однако стоит учесть, что:

- система FineReader, предназначенная для работы со сканированной информацией, не обеспечивает нужный темп и гибкость работы с текстами, напечатанными на бумаге, и совершенно не приспособлена к озвучиванию текстов, отображаемых на экране компьютерного монитора и других недоступных сканированию поверхностях;
- система JAWS не работает с печатными текстами и текстами, представленными в графических форматах в виде рисунка;
- звуковые книги требуют предварительной аудиозаписи речи в студии и её последующей трудоёмкой разметки.

Учитывая указанные ограничения, технология озвучивания плоскопечатных и компьютерных текстов, работающая в режиме, задаваемом самим пользователем, поддерживает возможности, не реализованные в настоящее время ни в одной из существующих программных систем.

Из известных аппаратных решений, близких по тематике, на рынке присутствует IRISPen. Это устройство — маркер по форме — подключается к компьютеру через USB интерфейс. С одной стороны устройства расположена камера. Предполагается, что пользователь ведёт маркер по строчке, в процессе его движения отсканированный текст транслируется на выбранный язык встроенной системой распознавания и перевода. Компания позиционирует устройство как предмет для перевода цитат информации с визитных карточек, т.е. работа с большими массивами данных не предусматривается. Очевидно, для использования этой системы необходимо видеть строки, поэтому такой вариант неприемлем для людей с нарушениями зрения.

Аналогичным по функциональности является устройство Intel Reader. Однако оно обладает сравнительно высокой стоимостью, на данный момент не поддерживает русский язык и позиционируется, скорее, как разработка, позволяющая реализовать чтение книжек вслух для людей без ограничений по зрению, что в свою очередь сказывается на управлении прибором.

## Основные компоненты технологии распознавания

Рассматриваемая технология распознавания и озвучивания текстов представлена на рис. 1. Она содержит три основных компонента:

- предварительную обработку изображения;
- распознавание символов;
- озвучивание распознанного текста.

Предварительная обработка изображения включает:

- ввод изображения в одном из стандартных графических форматов;
- преобразование изображения к монохромному представлению;
- определение контуров рисунка (преобразование к бинарному представлению);
- распознавание границ текстовых строк;
- определение прямоугольных фрагментов изображения, занимаемых символами в строке, включая составление списка их геометрических характеристик.

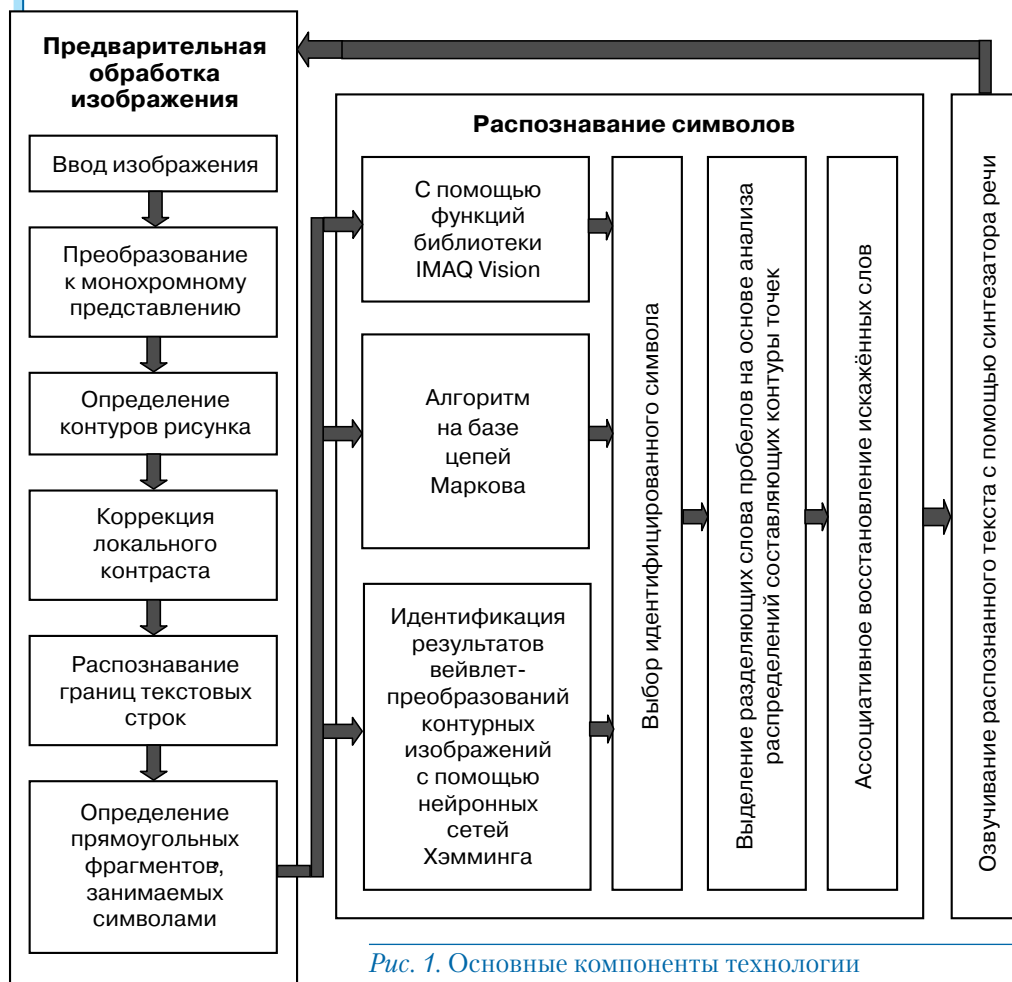


Рис. 1. Основные компоненты технологии распознавания и озвучивания плоскопечатных текстов

Для повышения надёжности распознавание символов производится независимо тремя различными способами:

- с помощью функций для обработки изображений из библиотеки IMAQ Vision, входящей в среду графического программирования LabVIEW [5] (этот встроенный набор функций в основном ориентирован на распознавание символов, включённых в штрихкоды, что делает механизм недостаточно эффективным при использовании на изображениях с низким разрешением и высоким уровнем помех);
- с применением нового алгоритма, использующего возможности цепей Маркова;
- путём идентификации результатов вейвлет-преобразований контурных символьных изображений с помощью релаксационных нейронных сетей Хэмминга [2–4]. Этот метод показал наивысшую эффективность при работе с сильно зашумлёнными изображениями, что связано как с особенностью вейвлет-преобразований, так и с вероятностной природой релаксационных нейронных сетей.

Подобное голосование уменьшает вероятность ложного распознавания. Символы, которые будут признаны нераспознанными, позже восстановит специальный механизм, опирающийся на словарь, что является стандартным методом в подобном случае.

Важными подзадачами, решаемыми в процессе распознавания текстов, являются:

- выделение пробелов;
- ассоциативное восстановление слов.

Выделение разделяющих слова пробелов происходит на основе анализа распределений, составляющих контуры точек. Восстанавливать искажённые слова приходится вследствие ошибок при идентификации символов.

### Особенности программной реализации и практического использования

Программная реализация выполнена в среде графического программирования LabVIEW. Использовались стандартные виртуальные инструменты для анализа данных, функции для обработки изображений из библиотеки IMAQ Vision, а также ряд процедур обработки изображений, выполненных в среде Borland Delphi и интегрированных в LabVIEW в форме динамически подключаемых библиотек.

Стандартный вариант применения рассматриваемой технологии предполагает сканирование текста с помощью веб-камеры, инициализация которой происходит в лучшем из доступных режимов<sup>1</sup>. Использование веб-камеры в качестве универсального считывающего устройства обусловлено её доступностью, компактностью, простотой и однотипностью в управлении на программном уровне. Для работы со всеми мультимедийными устройствами в Windows системах используется интерфейс программирования приложений DirectShow. Интерфейс USB камер поддерживается на большей части портативных компьютеров. В нашей жизни всё большее распространение получают т.н. субноуты. Они обладают низким энергопотреблением, которое обеспечивает большее время работы, и маленькими габаритами, что делает их ещё более мобильными по сравнению с другими портативными компьютерами. Совершенно очевидно, что их технические характеристики полностью удовлетворяют требованиям рассматриваемой системы, что делает доступной простому пользователю.

Для установления связи с веб-камерой используется DLL, работающий через DirectShow. В DirectShow включены функции активации устройства, функция запуска его в предпочтительном режиме, функция захвата изображения и функция завершения сеанса работы, что вполне достаточно для решения задачи. Аналогичная задача может быть решена с использованием компонентов IMAQ Vision, но в силу большей универсальности этот инструмент менее производителен, хотя опирается на абсолютно такие же принципы.

Введённое изображение преобразуется в монохромное чёрно-белое представление. Именно в таком виде наиболее удобны любые операции распознавания и анализа изображений. Для работы релаксационной нейронной сети символ переводится в контурный рисунок с помощью непрореженного вейвлет-преобразования на базе биортогонального вейвлета (рис. 2).

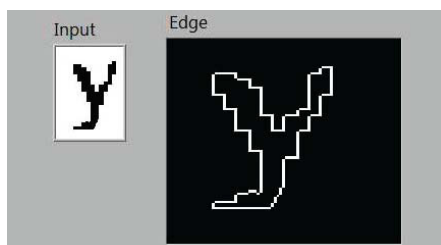


Рис. 2. Преобразование монохромного представления символа в контурный рисунок

<sup>1</sup> Как правило, распознавание проводилось с разрешением 320x240 и глубиной цвета 24bit.

Важной операцией, необходимой для корректной реконструкции связанного текста, является выделение из изображения символьных строк. Задача усложняется тем, что камера, как правило, удерживается в руке и, вследствие этого, строки попадают в кадр повёрнутыми. Для повышения надёжности распознавания при больших углах поворота необходимо восстанавливать горизонтальное положение строк. Для этого на основе анализа плотностей чёрных и белых точек, на нескольких вертикальных срезах изображения определяют местоположение начал и концов строк. Далее вычисляется средняя разность между значениями на срезах. Таким образом, мы получаем среднее расстояние, на которое конец строки смещён относительно начальной позиции. Последний подготовительный шаг — вычисление расстояния в горизонтальной плоскости. Исходя из полученных данных, возможно построить прямоугольный треугольник с известными сторонами и определить углы между катетами и гипотенузой. Угол  $\alpha$  на рис. 3 является искомым, после его вычисления изображение поворачивается (рис. 3, 4).

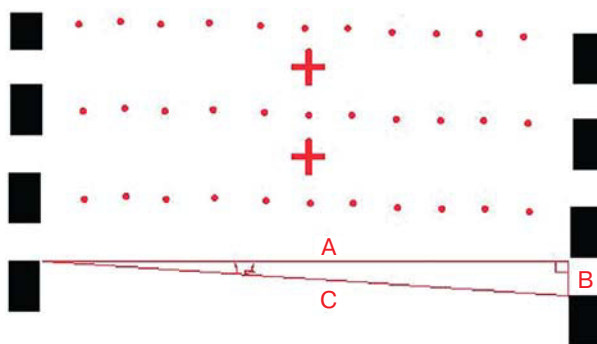


Рис. 3. Оценка расположения границ и угла поворота текстовых строк

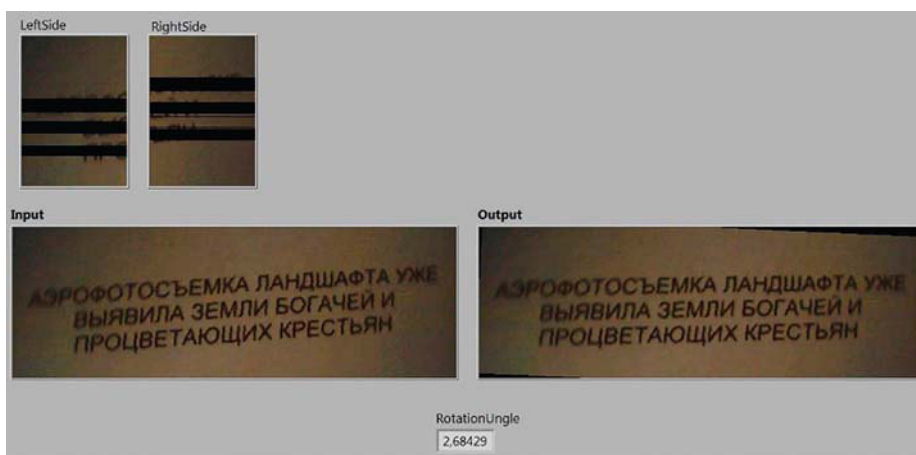


Рис. 4. Поворот тестовых строк

Поскольку для всех вычислений используются средние значения, метод не позволяет сделать строки параллельными границам изображения, но для работы остальных алгоритмов результат вполне приемлем. Ещё одним минусом подобной технологии является то, что в кадр обязательно должно попасть минимум две строки. Если это условие не выполняется,

изображение остаётся неизменным. Стоит отметить, что если строка обрывается на середине, это не влияет на вычисление в целом в силу того, что срезы делаются более чем на двух участках изображения. После такого поворота границы текстовых строк определяются повторно. На основе информации об этих границах фиксируются занимаемые символами прямоугольные фрагменты изображения (рис. 5).

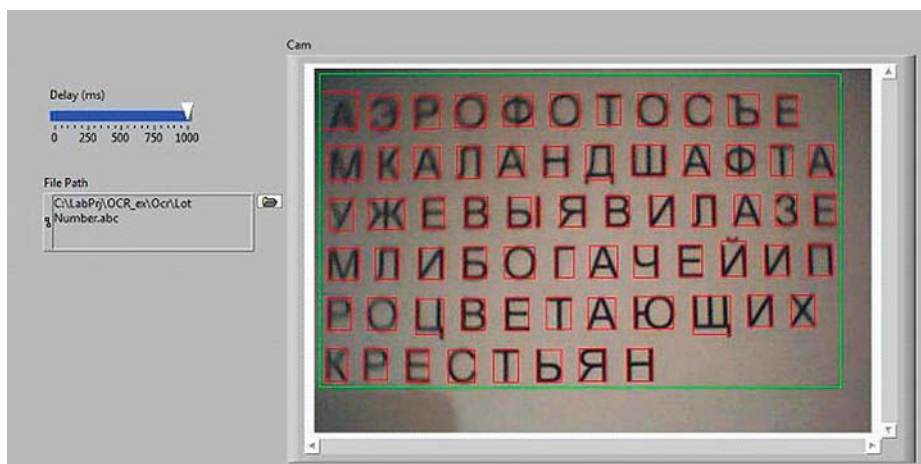


Рис. 5. Фиксация прямоугольных фрагментов изображения, занимаемых символами

По окончании выделения указанных прямоугольных фрагментов для всех символов в строке, соответствующий ей участок вырезается из кадра, просматриваемого веб-камерой, при этом малые фрагменты, не превышающие заданный порог, удаляются из строки. Оставшиеся фрагменты рассматриваются как области, содержащие распознаваемые символы произвольного размера. Указанная процедура повторяется для всех строк из кадра, просматриваемого веб-камерой. Информация о геометрических характеристиках строк и прямоугольных фрагментов изображения, включающих в себя символы, сохраняется для последующего распознавания в порядке расположения фрагментов в кадре.

Ослабить искажения в анализируемой части изображения позволяет алгоритм коррекции локального контраста. Этот приём позволяет уменьшить эффект, возникающий в связи с неравномерным освещением. При переводе в чёрно-белый формат невозможно установить общий уровень контраста для всего изображения, так как либо часть символов расплывётся, и распознать их будет невозможно, либо некоторые символы будут закрашены чёрным цветом. Суть метода заключается в следующем: на выделенных прямоугольных фрагментах перебираются все чёрные точки, из числа которых удаляются слабосвязанные (имеющие малое число одноцветных соседей). В то же время области, в которых предполагается исчезновение чёрных точек из-за нарушения освещённости или иных условий сканирования, заливаются чёрным цветом<sup>2</sup>. Указанный приём улучшает качество распознавания примерно на 10%, практически не влияя на скорость работы программы.

Изображение, вводимое с веб-камеры, обычно затемнено с одного или нескольких краёв, и несмотря на все использованные алгоритмы, с таким уровнем затемнения бороться бессмысленно. Эффект обусловлен непараллельностью поверхности текста и линзы, а также оптическими искажениями и недостаточным освещением. В случае значительных искажений анализируется только центральная часть графического представления.

<sup>2</sup> Примером может служить одна белая точка, все соседи которой в некоторой окрестности являются чёрными.

Один из реализованных способов распознавания символов опирается на возможности функций для обработки изображений, входящих в библиотеку IMAQ Vision среды графического программирования LabVIEW. После предварительного обучения на образцах пользователя он позволяет проводить идентификацию содержимого, сформированной ранее последовательности, прямоугольных фрагментов изображения. Результатом анализа является строка текста без пробелов, с обозначенными нераспознанными позициями.

Для обеспечения надёжности распознавания текста в особенно неблагоприятных условиях, характерных для рассматриваемой задачи, потребовалась реализация двух дополнительных независимых способов распознавания.

Первый опирается на возможности цепей Маркова.

Каждый прямоугольник с символом переводится в контурный рисунок с помощью непрореженного вейвлет-преобразования на базе биортогонального вейвлета, прямоугольные фрагменты с контурными изображениями последовательно преобразуются в два числовых вектора. Один из них описывает распределение точек контура вдоль оси абсцисс, другой — вдоль оси ординат. Таким образом, теряя часть информации, получаем простой инвариант длиной в 64 числа (символы отмасштабированы для окна 32x32). Такой вектор может быть проанализирован с удовлетворяющей для задачи скоростью. Полученный вектор подвергается быстрому вейвлет-преобразованию с использованием вейвлетов Добеши 4-го порядка. Полученные подмножества последовательностей вейвлет-коэффициентов  $\mathbf{W}=\{\mathbf{w}_{i,j}\}_{i,j=1,\dots,N}$  рассматриваются как представления символов и анализируются с помощью распознающей цепи Маркова, структура которой представлена на рис. 6. Процедура распознавания основана на анализе знаков вейвлет-коэффициентов из указанной последовательности.

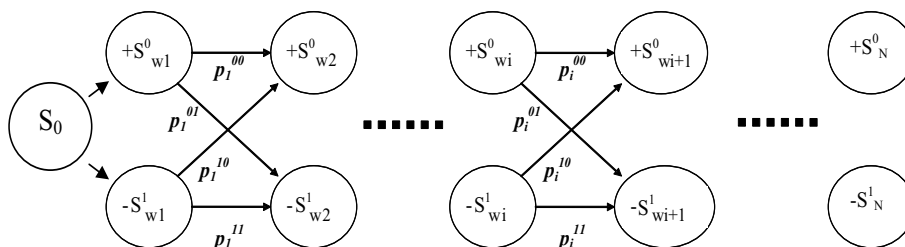


Рис. 6. Структура цепи Маркова

Система, описываемая данной цепью, функционирует  $N$  моментов времени, где  $N$  — длина анализируемой последовательности вейвлет-коэффициентов. В начальный момент времени система находится в состоянии  $S_0$ . В  $i$ -й момент времени, где  $i$  изменяется от 1 до  $N$ , система может быть или в состоянии  $+S_i^0$ , или в состоянии  $-S_i^1$ , а именно: в состоянии  $+S_i^0$  — в случае неотрицательности  $i$ -го элемента последовательности  $\mathbf{W}$  и в состоянии  $-S_i^1$  — в случае его отрицательности. Система является нестационарной: изменяется матрица  $\mathbf{P}_i=\|p_{i,j}\|$  вероятностей переходов между состояниями, такая, что  $\mathbf{p}_i=\mathbf{P}_i\mathbf{p}_{i-1}$ , где  $\mathbf{p}_i$  — вектор-столбец вероятностей пребывания в состояниях цепи в  $i$ -й момент времени, причём в указанный момент времени все её элементы равны нулю, кроме вероятностей переходов между парами состояний  $S_{i-1}^1$  и  $S_i^0$ ,  $S_{i-1}^0$  и  $S_i^1$ ,  $S_{i-1}^0$  и  $S_i^0$ ,  $S_{i-1}^1$  и  $S_i^1$ , которые далее обозначены, соответственно, как  $p_i^{11}$ ,  $p_i^{01}$ ,  $p_i^{10}$  и  $p_i^{00}$ . В каждый из моментов времени  $i$  выполняются следующие нормировочные условия:  $p_i^{11}+p_i^{10}=1$  и  $p_i^{01}+p_i^{00}=1$ .



Последовательности  $W_a$ , соответствующей заданному символу  $a$ , приведённые правила ставят в соответствие определённую последовательность  $S_a$  прохождения состояний рассматриваемой цепи Маркова<sup>3</sup>.

При обучении цепи Маркова на алфавите  $A=\{a_m\}$  для каждого символа  $a_m$  формируется множество его эталонов<sup>4</sup>  $a_{mk}$ ,  $k=1, \dots, K_m$ . Каждому символу  $a_m$  ставится в соответствие цепь Маркова, вероятности переходов, между состояниями которой определяются методом максимального правдоподобия. Для этого вычисляется распределение вероятностей переходов, обеспечивающее наибольшую вероятность  $P$  появления заданного множества последовательностей состояний, соответствующих эталонам данного символа:  $P=\prod_{k=1, \dots, K_m} P_k$ , где  $P_k=\prod_{i=1, \dots, N} p_i^{signWi-1, k \ signWi, k}$ , где  $k$  — номер эталона символа, а  $i$  — номер состояния, причём полагается, что  $signw_{0, k}=1$ .

Указанная задача может быть решена одним из подходящих численных методов нелинейного программирования<sup>5</sup>. С целью упрощения решения удобно перейти к критерию  $\ln P$ , монотонно связанному с исходным критерием  $P$ , переформулировав исходную постановку в задачу оптимизации для величин  $\ln p_i^{signWi-1, k \ signWi, k}$ , на которые наложено условие неположительности<sup>6</sup>.

Распознавание поступающих на вход системы новых символов возможно после завершения обучения. При этом для цепи Маркова, соответствующей каждому из символов алфавита  $A$ , вычисляется вероятность прохождения последовательности состояний, определяемой вейвлет-коэффициентами  $\{w_i\}_{i=1, \dots, N}$  распознаваемого символа, которая равна  $\prod_{i=1, \dots, N} p_i^{signWi-1 \ signWi}$ . Тип символа, который соответствует цепи Маркова, имеющей наибольшую вероятность такого рода, выдаётся как результат распознавания.

Оценка вероятности корректного распознавания символов с помощью описанного алгоритма, полученная по результатам 1000 проб на базе асимптотической аппроксимации биномиального распределения, равна 0,791, причём нижняя и верхняя границы 95% доверительного интервала есть 0,76 и 0,83.

Оказалось, что достаточно использовать последовательности  $W$ , состоящие из двенадцати вейвлет-коэффициентов ( $N = 12$ ), начиная с третьего коэффициента разложения распределений контурных точек. Расчёты выявили неинформативность первых двух вейвлет-коэффициентов этого разложения для решения задачи распознавания.

Второй способ реализует новый алгоритм [2–4], использующий возможности вейвлет-преобразований и релаксационных нейронных сетей. Он способен эффективно работать после обучения на ограниченном числе образцов (рис. 7 стр. 88).

В этом методе используется сеть Хэмминга. Она требует значительно меньших затрат по сравнению с сетью Хопфилда, часто используемой в подобных случаях. При решении поставленной задачи нет необходимости восстанавливать образец по входам сети, т.е. нужен только номер предполагаемого образца. Таким образом, вместо классического варианта с ассоциативной памятью выбран алгоритм с релаксационной нейронной сетью. Суть данной структуры заключается в поиске хэммингова расстояния<sup>7</sup> от представленного образца до образца из обучающей выборки. Образец, до которого такое расстояние окажется наименьшим, признаётся искомым. Так же характерным признаком сети является то, что нейроны её выходного слоя связаны между собой дополнительными отрицательными обратными синаптическими связями. Единственный синапс, с положительной обратной связью для каждого нейрона,

<sup>3</sup> Следует заметить, что рассматриваемая цепь Маркова может быть описана и в терминах конечного автомата.

<sup>4</sup> Эталоны, как правило, соответствуют различным вариантам изображения символов, принятым в разных шрифтах.

<sup>5</sup> Для этого, в частности, может быть использован модуль Solver электронной таблицы Microsoft Excel.

<sup>6</sup> Поскольку  $0 \leq p_i^{signWik \ signWi+1, k} \leq 1$  для всех  $i$  и  $k$ .

<sup>7</sup> Расстояние Хэмминга – число отличающихся битов в двух бинарных векторах.

соединён с его же аксоном. Это позволяет сделать пространство поиска более «контрастным» и, как следствие, улучшить распознавание в сложных случаях.

На первом этапе нейросетевого метода распознавания прямоугольные фрагменты с контурными изображениями последовательно преобразуются в два числовых вектора. Один из них описывает распределение точек контура вдоль оси абсцисс, другой — вдоль оси ординат. Таким образом, теряя часть информации, получаем простой инвариант длиной в 64 числа (символы отмасштабированы для окна 32x32). Такой вектор может быть проанализирован с удовлетворяющей для задачи скоростью. Полученный вектор подвергается быстрому вейвлет-преобразованию, результаты которого подаются на вход сети Хэмминга с радиальными базисными элементами и экспоненциальными функциями активации [2–4]. Весовые коэффициенты радиальных элементов этих сетей вычисляются в соответствии с имеющимися эталонными образцами распознаваемых символов. После циклических вычислений нейронная сеть сходится к номеру ближайшего эталона. Таким образом, сходство с одним из заданных эталонных символов определяется нахождением в соответствующей области притяжения в пространстве допустимых представлений входного фрагмента. Последовательность обработки данных в этом методе распознавания представлена на рис. 7.

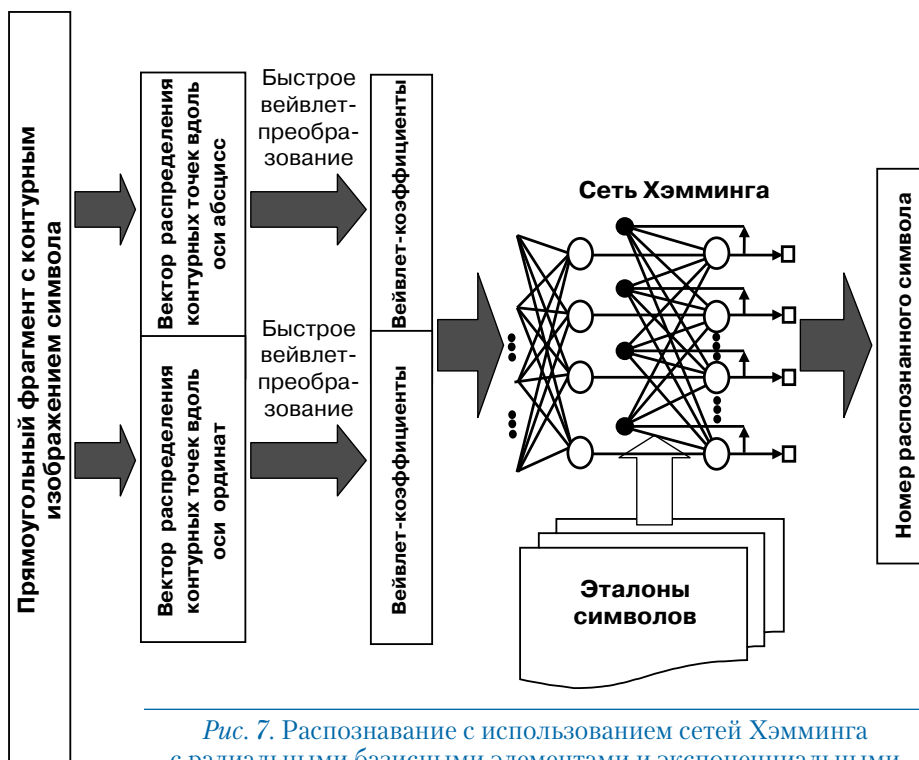


Рис. 7. Распознавание с использованием сетей Хэмминга с радиальными базисными элементами и экспоненциальными функциями активации

Символ считается идентифицированным, если он выдаётся в качестве результата не менее чем двумя используемыми способами.

Если распознанная строка символов не будет содержать информацию о разделяющих слова пробелах, то синтезатор речи не сможет корректно воспроизвести полученный текст. Выявление пробелов в тексте необходимо

для обеспечения работы синтезатора речи. Поиск пробелов происходит на основании частотного анализа.

Расстояния между буквами обычно меньше, чем между словами, в полтора-два раза. Этот факт позволяет говорить о весьма устойчивом разделении слов текста. Данный анализ производится после нахождения прямоугольных фрагментов изображения, содержащих символы, поэтому можно легко определить позиции, после которых следует вставлять пробелы. Пробелы вставляются в текстовую строку перед заключительной коррекцией полученного текста, которая проводится на последнем этапе распознавания. Эта коррекция выполняется путём ассоциативного восстановления слов (искажённых вследствие ошибок при идентификации символов либо содержащих нераспознанные элементы) с помощью встроеного словаря.

Для того чтобы текст был озвучен, необходимо подготовить строку, содержащую последовательно все символы с пробелами. Эта строка передаётся заранее настроенному синтезатору речи. Помимо процедуры инициализации, синтезатор требует указать, какой из возможных вариантов голоса использовать при озвучке. Этот профиль, естественно, тоже можно выбрать в программной оболочке. По умолчанию используется один голос, профиль которого можно перевести как женский взрослый. Также программа осуществляет контроль за тем, завершился процесс воспроизведения или нет. Если процесс не завершился, новая порция текста не передаётся, так как передача вызовет заполнение буфера и воспроизведение длительной последовательности звуков уже после завершения сеанса работы с программой. Для озвучивания распознанного текста используется стандартный синтезатор речи. В среде операционной системы Windows для синтеза речи может быть использован интерфейс прикладного программирования Microsoft SAPI (Speech Application Programming Interface).

Очередной кадр изображения обрабатывается после озвучивания предыдущего. Темп обработки можно регулировать по желанию пользователя. Это вполне разумно, поскольку для восприятия различных по уровню информационной насыщенности текстов требуется, как правило, разное время. Не следует забывать и об индивидуальных различиях в скорости восприятия информации. Установка заниженного темпа обработки может быть полезна на начальных этапах работы с системой, пока пользователь не привык к новому способу восприятия информации, по сути требующему координированной работы рук, слуха и одновременной интерпретации услышанного.

Текущая версия программной реализации эффективно распознаёт и озвучивает произвольный плоскочечатный текст или текст на экране компьютерного монитора при достаточном освещении. Проведённое тестирование позволяет говорить о том, что не существует букв русского алфавита, плохо идентифицируемых алгоритмом.

## Основные результаты и выводы

Разработана и программно реализована технология обработки текстов, особенностями которой являются:

- интеграция в одном программном продукте средств сканирования, распознавания и озвучивания;
- наличие развитых средств устранения ошибок, обусловленных низким качеством сканированного изображения;
- применение нового алгоритма распознавания символов, опирающегося на возможности вейвлет-преобразований и релаксационных нейронных сетей и способного эффективно работать после обучения на ограниченном числе образцов;
- разработан и программно реализован новый алгоритм распознавания символов на базе цепей Маркова, даны формальные оценки его эффективности;
- использование веб-камеры для сканирования озвучиваемых изображений.



Разработанная технология имеет значимые преимущества перед существующими в настоящее время средствами озвучивания текстов для людей с нарушениями зрения, связанные с:

- мобильностью аппаратных средств;
- высокой скоростью и гибкостью воспроизведения информации в удобном для пользователя режиме;
- возможностью работы с текстами, представленными не в электронной форме;
- способностью работать с изображениями на экране компьютерного монитора.

Представленные средства могут быть использованы:

- для чтения литературы, изданной традиционным плоскочечатным способом;
- для работы за компьютером с обычным программным обеспечением, не предназначенным для незрячих пользователей.

## Литература

1. Богомолов А.М. Личностный адаптационный потенциал в контексте системного анализа. Психологическая наука и образование. 2008. № 1. С. 67–73.
2. Куравский Л.С., Баранов С.Н., Буланова О.Е., Кравчук Т.Е. Нейросетевая технология диагностики патологических состояний по аномалиям электроэнцефалограмм. Нейрокомпьютеры: разработка и применение. 2007. № 4. С. 4–14.
3. L.S. Kuravsky and S.N. Baranov. Wavelet transforms and relaxation neural networks as promising technology components of technical and medical diagnostics and monitoring. — In: Proc. 2nd World Congress on Engineering Asset Management and 4th International Conference on Condition Monitoring, Harrogate, United Kingdom, June 2007. Pp. 1117–1132.
4. L.S. Kuravsky and S.N. Baranov. Technical diagnostics and monitoring based on capabilities of wavelet transforms and relaxation neural networks. Insight. Vol. 50. No 3. March 2008. Pp. 127–132.
5. LabVIEW tutorial for Windows. National Instruments Corp., 2004–2007.
6. FineReader ([http://www.abbyy.ru/upload/files/FineReader\\_9.0\\_Reviewer%27s\\_Guide\\_Russian.pdf](http://www.abbyy.ru/upload/files/FineReader_9.0_Reviewer%27s_Guide_Russian.pdf)).
7. JAWS (<http://www.cardiff.ac.uk/accessibility/technicalinformation/guidetojaws/index.html>).
8. CuneiForm ([ftp://ftp.dol.ru/pub/users/cgntv/download/cuneiform/eng/cunei\\_e.pdf](ftp://ftp.dol.ru/pub/users/cgntv/download/cuneiform/eng/cunei_e.pdf)).
9. Adriane Knoppix (<http://www.knopper.net/knoppix-adriane/index-en.html>).

## Сведения об авторах

### **Юрьев Григорий Александрович** —

аспирант факультета информационных технологий Московского городского психолого-педагогического университета.  
nezdeshni@gmail.com

### **Куравский Лев Семёнович** —

декан факультета информационных технологий, заведующий кафедрой прикладной информатики МГППУ, профессор, доктор технических наук. kuravsky@yahoo.com