

УДК 004.056

Р.В. Мещеряков

Система оценки качества передаваемой речи

Описываются постановка задачи оценки качества речевого сигнала. Описывается алгоритмическое и программное обеспечение поддержки работы эксперта в соответствии со стандартом оценки качества речевого сигнала.

Ключевые слова: речевой сигнал, слух, система, автоматизированная система.

Введение

В ряде случаев необходимо оценивать качество речевого сигнала, в частности при передаче речевого сигнала по каналам связи, при врачебных исследованиях, при постановке певческого голоса. Как правило, это сложная операция, для которой собирается специальная аудиторская бригада. Однако средств автоматизации недостаточно для массового использования.

Целью работы являются создание и апробация автоматизированной системы, а также проведение оценки качества речевого сигнала с помощью созданного программного обеспечения. В качестве базового стандарта принят «ГОСТ 50840-95 Передача речи по трактам связи. Методы оценки качества, разборчивости и узнаваемости» [1]. Данный стандарт регламентирует получение комплексной оценки качества передачи речи, основанной на методах измерения показателей разборчивости и качества речи, а также узнаваемости голоса диктора.

Математическое обеспечение

Одной из наиболее важных характеристик речи является ее разборчивость – относительное количество (в процентах) правильно принятых элементов (слогов, слов, фраз) артикуляционных таблиц. Качество речи – величина, характеризующая субъективную оценку звучания речи в испытуемом тракте [1]:

- по сравнению со звучанием в контрольном тракте (принятым за 5 баллов):
- неразборчиво;
- едва разборчиво, различимы некоторые слова;
- разборчиво со значительными трудностями;
- разборчиво практически без затруднения;
- отличная разборчивость;
- по сравнению со звучанием речи в другом тракте (в процентах предпочтения).

Неправильная речь – это речь, которая затрудняет понимание. Это происходит и при использовании незнакомых слов, тем более при их употреблении не с тем значением, которое зафиксировано в словарях. Такое же затруднение вызывает использование жаргонных слов, особенно при их широком распространении; затрудняет понимание и, следовательно, противоречит критерию хорошей речи не свойственная русскому языку сочетаемость русских слов.

Как отмечалось ранее, измерения проводит бригада операторов, прошедшая специальное обучение (тренировку), путём прослушивания слоговых артикуляционных таблиц. Чтение слогов осуществляется диктором ровным голосом, чётко, но без подчёркивания отдельных звуков с постоянным уровнем речи, который контролируется шумомером. Слоги следует читать в следующем ритме: 1 слог в $(3 \pm 0,3)$ с. Диктор должен выдерживать постоянный ритм речи на протяжении чтения всей таблицы.

Дальше проводится цикл измерений, который включает в себя приём всеми аудиторами от всех дикторов по $5k$ таблиц, где $k = 1, 2, 3, \dots$. Пятёрки таблиц должны иметь номера 1–5, 6–10, 11–15 и т.д. Вычисляется среднее значение разборчивости (S) по формуле

$$S = \frac{1}{N} \sum_{i=1}^N S_i, \quad (1)$$

где S_i – результат единичного измерения, % (диктор–таблица–аудитор); N – число единичных измерений.

Далее выявляют сомнительные значения S_i , которые отбрасывают, и вычисляют новое значение S . Время работы бригады должно быть не более 4 ч за один день. После приёма 5 таблиц делается перерыв 5–10 мин. Общее число таблиц за одно измерение – до 40.

При работе в акустических шумах бригада приступает к измерениям спустя 5–10 мин после пребывания в условиях шума. Общее число таблиц – 30 (при уровне шума 80–100 дБ) и 20 (при уровне шума более 100 дБ). Метод следует использовать при аттестации тракта связи.

При наличии дополнительного оборудования (ПЭВМ с устройством ввода-вывода речевой информации) измеряют разборчивость методом дописывания неполных слогов. В слогах таблицы отсутствуют начальные (мягкие/твердые) или конечные согласные. Аудиторы прослушивают таблицы, выводимые из ПЭВМ, и одновременно наблюдают на экране дисплея неполные слоги. По данным артикуляционных измерений вычисляют среднюю разборчивость (Q) по формуле

$$Q = \frac{1}{N} \sum_{i=1}^N q_i, \quad (1)$$

где q_i – результат единичного измерения, % (диктор–таблица–аудитор); N – число единичных измерений.

Метод дописывания рекомендуется для оценки трактов в условиях низкой разборчивости, например в шумах. При этом достигаются автоматизация процедур получения, обработки и углубленного анализа данных артикуляционных измерений, ускорение получения результатов измерений.

Измерение качества речи методом парных сравнений с контрольным трактом проводят путем прослушивания пар одинаковых фраз длительностью по 2–3 с каждая. Аудитор проставляет в бланке свои оценки качества речи по пятибалльной шкале с точностью до 0,1. Одинаковые оценки не допускаются. Обработку полученных результатов проводят путём вычисления средней балльной оценки испытуемого и контрольного трактов по формуле

$$X_{\infty(k)i} = S = \frac{1}{N} X_{u(x)i}, \quad (3)$$

где $X_{u(x)i}$ – результат единичного измерения, % (диктор–таблица–аудитор); N – число единичных измерений.

Далее проводят нормативные оценки испытуемого тракта X_u по формуле

$$X = \frac{X_u}{X_k} \times 5. \quad (4)$$

Метод рекомендуется использовать при аттестации аппаратуры по показателю качества речи.

Измерения качества речи методом оценки по селективным признакам проводят путём прослушивания аудиторами фраз. Фразы для прослушивания передают с интервалам 2–3 с. Число прослушиваний каждой фразы не ограничено. Аудиторы проводят сравнение звучания фразы, прошедшей через контрольный тракт, и фразы, прошедшей через испытуемый тракт (аппаратуру), и определяют наличие следующих селективных признаков искажения в звучании речи относительно контрольного тракта:

- картавость;
- плаксивость
- гнусавость;
- механический голос;
- дребезжание, хрип;
- помеха в паузах речи.

Кроме того, проводятся измерения разборчивости, качества речи и узнаваемости голоса диктора методом парных условий, фазовой разборчивости при ускоренном темпе произнесения.

Программное обеспечение

Программное обеспечение реализовано П.С. Шороховым в виде оконного приложения и предназначено для выполнения всех операций, предусмотренных [1]. Опишем основные функции и экранные формы (рис. 1).

В верхнем левом углу выбираем «Для диктора». Появляется вкладка «Создать тест». Нажимаем, появляется новое диалоговое окно «Создание теста. Шаг 1». На нём видим методы, по которым будем проводить оценку качества (рис. 2).

Выбираем первый метод «Измерение разборчивости речи артикуляционным методом» и жмём далее (рис. 3).

На форме имеется два окна. Слева в окне видим список файлов. Справа пустое окно. Если нажимаем на файл в левом окне, то в правом окне появляется его содержимое. Снизу видим путь, где располагается эта таблица (рис. 4).

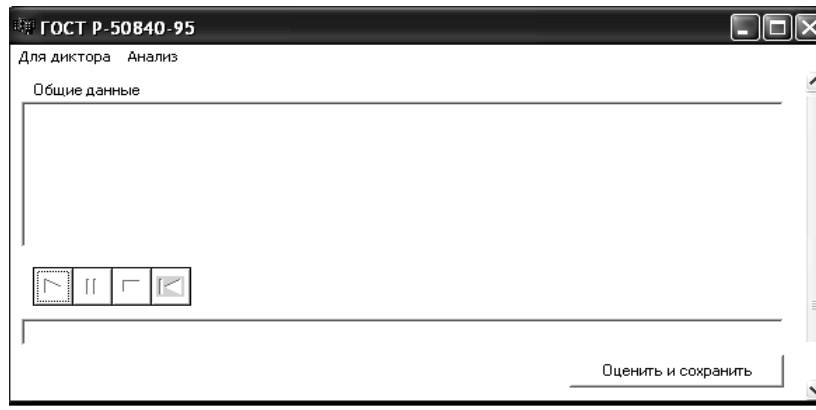


Рис. 1. Главный вид программы

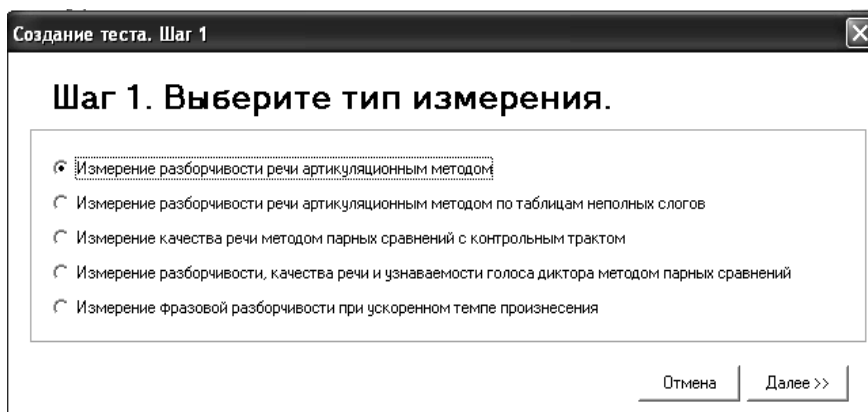


Рис. 2. Создание теста. Шаг 1

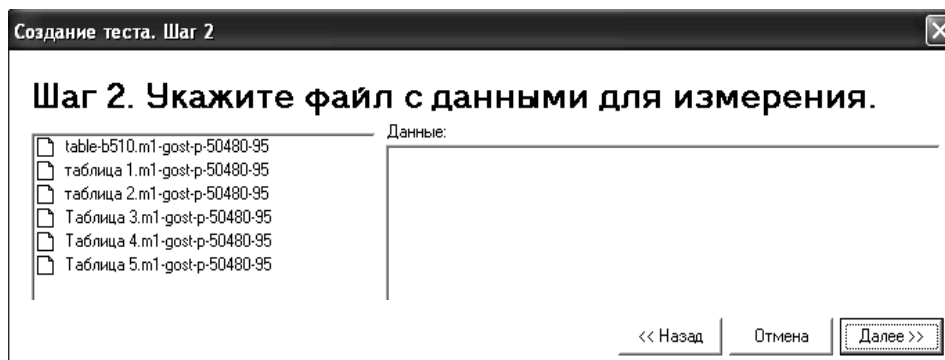


Рис. 3. Создание теста. Шаг 2

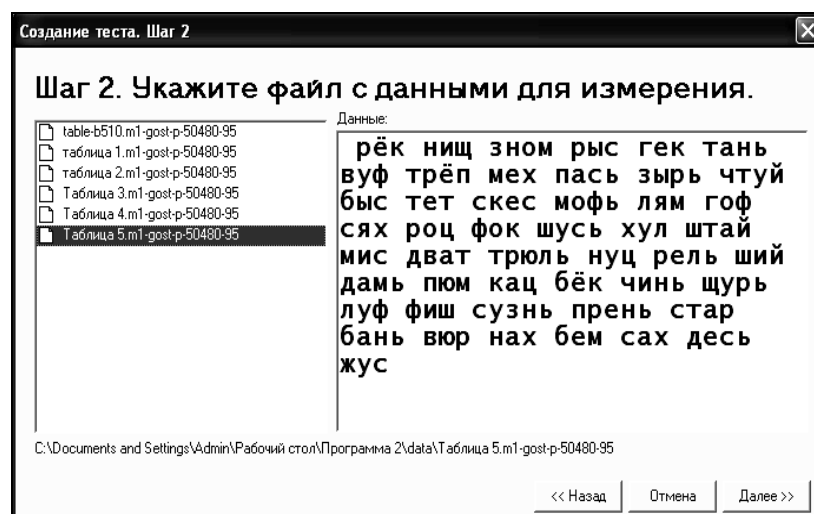


Рис. 4. Создание теста. Шаг 2

Нажимаем «далее», появляется форма «Создание теста». Шаг 3 (рис. 5).

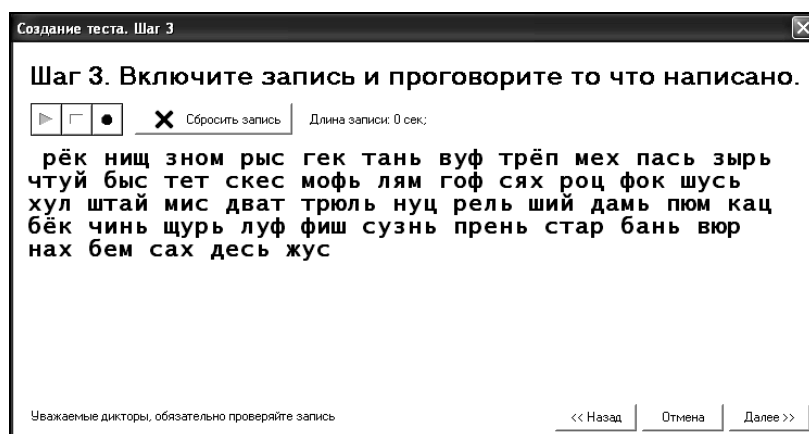


Рис. 5. Создание теста. Шаг 3

Включаем запись и проговариваем то, что написано на форме. Слоги следует читать в следующем ритме: 1 слог в 3 с. После того как прочитали и записали слог можем прослушать то, что получилось путём нажатия на кнопку «Воспроизвести». Нажимаем далее (рис. 6.)

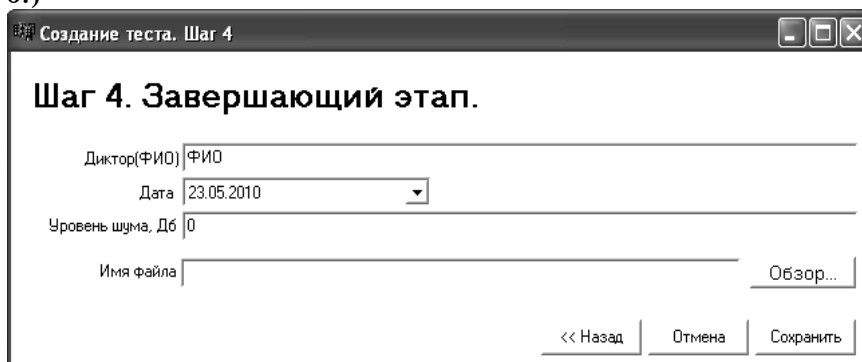


Рис. 6. Создание теста. Шаг 4

В поле «Диктор (ФИО)» записываем ФИО диктора, который читал таблицу.

В поле «Дата» вносим число, когда была записана таблица.

В поле «Уровень шума, дБ» записываем, при каком уровне шума читалась таблица. (Примечание: от уровня шума зависит, какое количество таблиц может прочитать диктор).

В поле «Имя файла» выбираем, куда сохранить файл и под каким названием. Сохраненный файл перетаскиваем на главную форму. На ней видим в верхнем окне в поле «Общие данные», какой метод использовался, кто был диктором, дату записи и уровень шума.

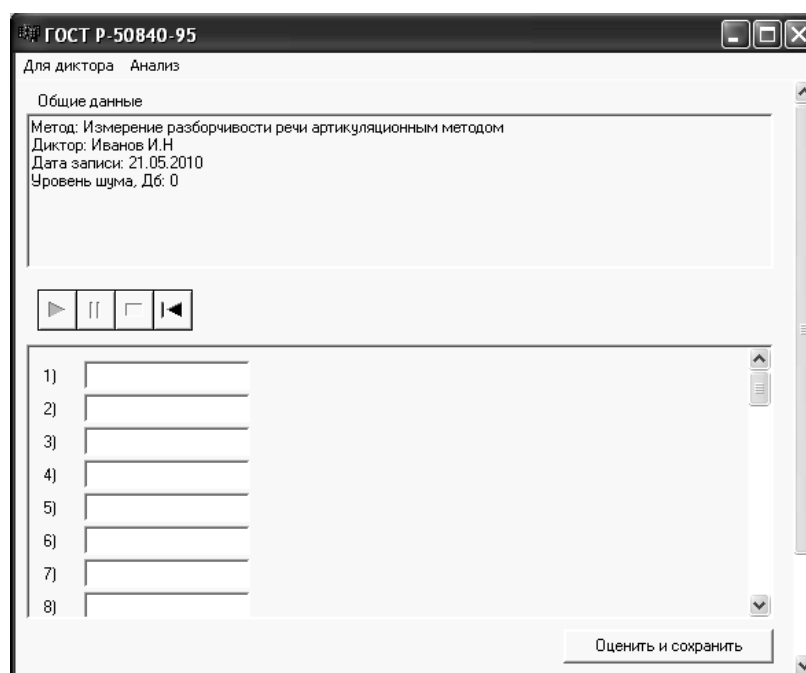


Рис. 7. Главная форма для аудитора

В нижнем окне появились строки от 1–50 (количество слогов из таблицы). Нажимаем кнопку «Проигрывать». То, что слышим, записываем в эти строки (рис. 8).

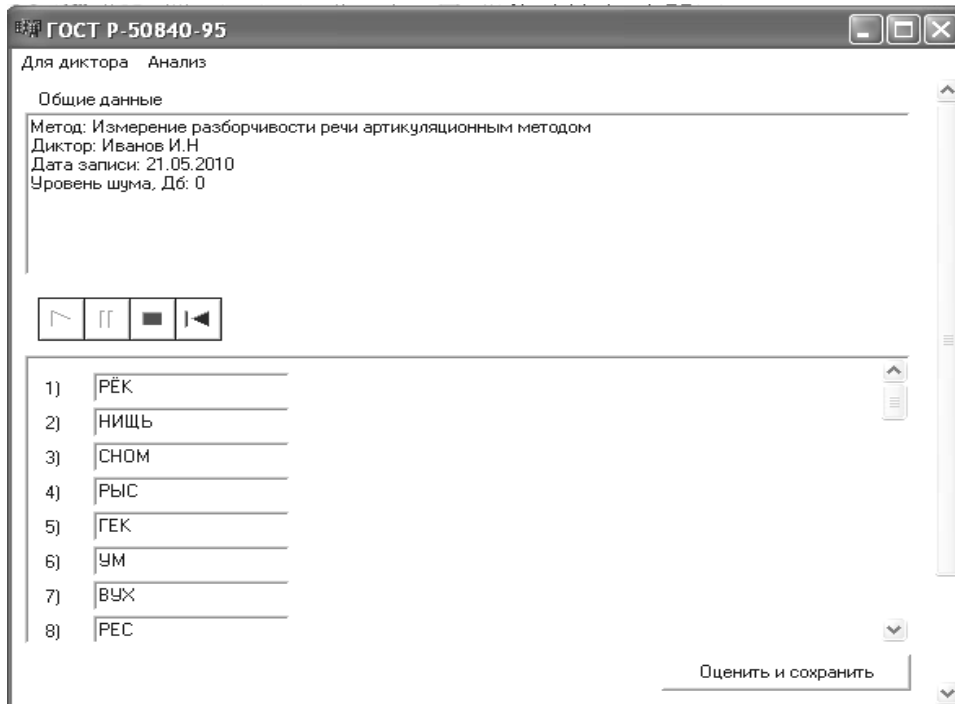


Рис. 8. Главная форма с заполненными строками

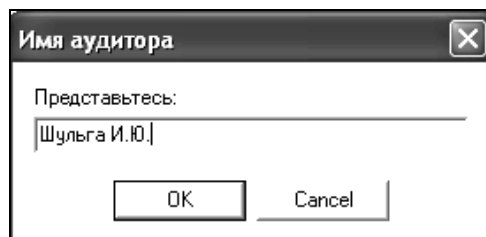


Рис. 9. Имя аудитора

Дальше нажимаем кнопку «Оценить и сохранить» появляется окно, в котором указываем ФИО аудитора, который проводил оценку (рис. 9).

Дальше нажимаем «Анализ». В появившемся окне добавляем наш файл и рассчитываем среднюю оценку (рис. 10, 11).

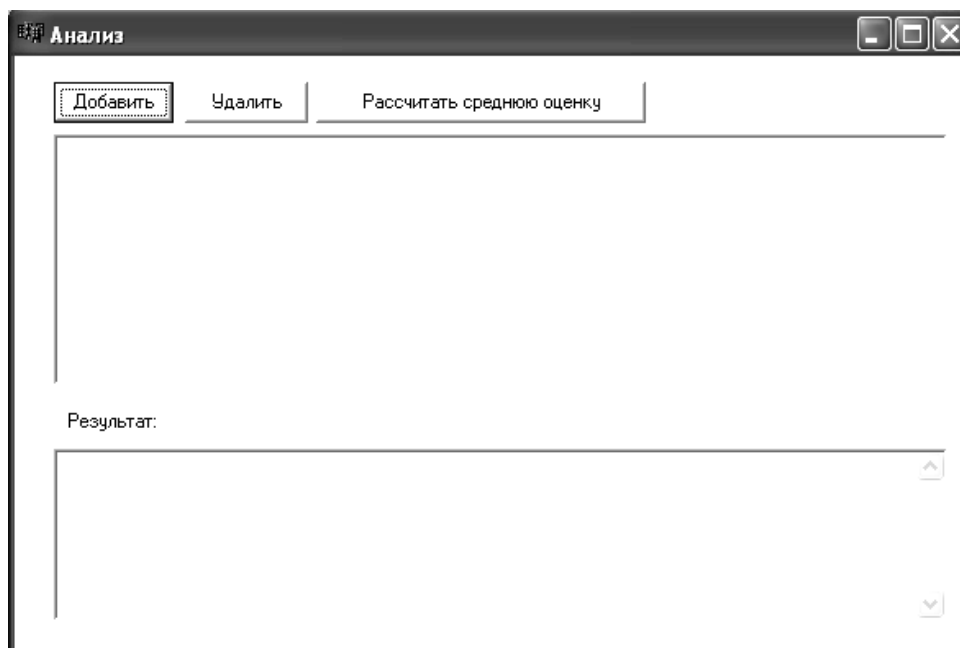


Рис. 10. Форма анализа

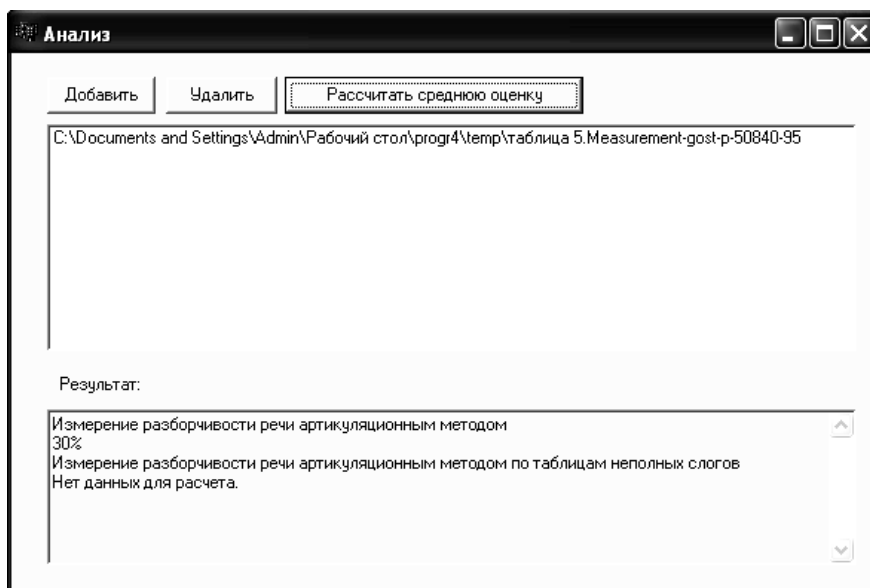


Рис. 11. Результат

Заключение

Предложенный подход к автоматизированной оценке качества речевого сигнала с одной стороны, позволяет использовать стандартизированные способы оценки согласно [1]. С другой стороны, это позволяет повысить скорость проведения этих исследований, а также заложить возможность интегрированных оценок речевого сигнала [2], использование которых делает возможным формирование итоговой технико-лингвистической оценки качества речевого сигнала.

Работа выполнена в рамках реализации программы «Научные и научно-педагогические кадры инновационной России на 2009–2013 годы» (контракт № П1083).

Литература

1. ГОСТ Р 50840-95. Передача речи по трактам связи. Методы оценки качества, разборчивости и узнаваемости. – М.: Госстандарт России, 1995. – 234 с.
2. Мещеряков Р.В. Использование информационных критериев для оценки иерархических диалоговых систем // Научно-технические ведомости СПбГПУ. – 2009. – № 4(82). – С. 113–122.

Мещеряков Роман Валерьевич

Канд. техн. наук, доцент каф. комплексной информационной безопасности
электронно-вычислительных систем ТУСУРа
Тел.: (382-2) 41-34-26
Эл. почта: mrv@keva.tusur.ru

Mescheriakov R.V.

System of the transmitted speech quality estimation

The problem of speech signal quality is stated. Algorithms and software for expert work support, in accordance with the standard of the speech signal quality estimation, are described.

Keywords: Speech signal, hearing, the system, the automated system.