

**Оценка возможности применения программного анализа
устной речи для исследований по разработке метода
диагностики наличия у человека скрываемой им информации**
06, июнь 2015

DOI: 10.7463/0615.0780454

Холодный Ю. И.¹, Панин С. Д.¹, Белых В. А.^{2,*}

УДК: 159.96

¹Россия, МГТУ им. Н.Э. Баумана

²Центральный научно-исследовательский радиотехнический институт
им. академика А.И. Берга

* v.a.belykh@gmail.com

Введение

Из повседневной жизни хорошо известно, что многих людей, когда они пытаются утаить что-то от собеседника и сообщают ему заведомо ложную информацию, их выдаёт собственная речь – меняется громкость и тембр голоса, появляются ненужные, логически необоснованные повторы речи, неожиданно возникают паузы в речи, изменяется её скорость и т.д.

Очевидно, что в некоторых областях общественной практики (раскрытие и расследование преступлений, профотбор, контроль работодателем деятельности персонала и др.) существует потребность в выявлении (диагностике) у человека наличия у него скрываемой информации о каких-то событиях прошлого и, как следствие, – в технических средствах для решения указанных задач.

Принято полагать, что история технических средств для целей выявления у человека скрываемой информации по его голосу началась в 1970 г., когда в США был создан первый прибор, названный «Psychological stress evaluator» («PSE-1»). В отечественной специальной литературе приборы этого класса получили наименование «анализаторы стресса по голосу» (далее – АСПГ) от английского названия «voice stress analyzer (VSA)».

В последующие годы разными фирмами были созданы АСПГ – «PSE-101», «MARK», «MARK-II», «CVSA», «CVSA-II» /США/; «Truster Pro» /Израиль/ и ряд других. (В частности, две модификации АСПГ израильского производства продаются на российском рынке одной из коммерческих фирм).

Фирмы-производители АСПГ активно рекламируют их прикладные возможности, что нередко оказывает внушающее воздействие на различных потребителей этой рекламы. Примером тому может служить статья молодых исследователей МГТУ им. Н.Э. Баумана,

которые, будучи недостаточно осведомленными в области теории и практики применения АСПГ, без проведения каких-либо экспериментальных исследований увлеченно утверждали, что «CVSA II эффективен во всех следственных ситуациях, таких как: убийство, сексуальные преступления, ограбления, должностные преступления и внутренние расследования, а также при проверках перед приемом на работу в правоохранительные органы» [1].

Но на протяжении всей истории существования АСПГ между их сторонниками и противниками идет спор о прикладной эффективности этих приборов. Многочисленные экспериментальные исследования, выполненные в лабораторных и полевых условиях, показали низкую достоверность получаемых с их помощью результатов [2].

В советской период отечественная психологическая и юридическая наука констатировали, что «определение факта ложности показаний и выявление истины из «подтекста» ложного сообщения являются проблемами чрезвычайно сложными и пока далекими от разрешения. Проблема определения лжи, заключенной в тексте, является полностью проблемой психолингвистической» [3].

На рубеже 1970-1980-х годов один из авторов данной публикации провел экспериментальной изучение возможности использования психолингвистических характеристик речи для диагностики у человека скрываемой им информации. Позднее, в результате совместных экспериментальных работ с группой специалистов под руководством проф. Носенко Э.Л., в речи были выделены несколько каналов «утечки информации» от человека, каждый из которых оказался полезным для диагностики «ложных сообщений». Этими каналами оказались – амплитудно-частотные, темпоральные, лексические и синтаксические характеристики устной речи. В частности, полученный в ходе исследования материал был оценен «по ряду лингвистических и паралингвистических показателей, характеризующих: а) степень наличия в предложении различных структурных элементов (существительных, прилагательных, глаголов, наречий), дающую представление о «динамичности» речи; б) разнообразие строевых слов (предлогов, союзов, местоимений и т.д.), дающее представление о степени расчлененности и сложности синтаксической организации речи; в) словарное разнообразие, свидетельствующее о степени повторяемости лексических единиц текста; г) интенсивность различного рода «явлений нерешительности», прерывающих поток речи и временной организации высказывания, сигнализирующие о степени затрудненности процесса выражения мыслей» и проч.» При этом было установлено, что «по паралингвистическим и лексическим параметрам различия были статистически значимы» [4].

Исследования показали высокую эффективность избранных характеристик и параметров для решения экспериментальных задач, но в 1991 г. работы в этом направлении были прекращены и более не возобновлялись^{*)}.

^{*)} Экспериментальные исследования проводились совместно с Днепропетровским государственным университетом и были прекращены в связи с распадом СССР.

По-видимому, единственным в постсоветский период ученым, который обратил внимание на возможность использования речи человека в целях выявления у человека скрываемой информации и обнаружения в его памяти следов уголовно релевантных событий, явился украинский криминалист М.В. Салтевский, посвятивший этому отдельную главу («Сліди пам'яті людини») в учебнике по криминалистике [5]. В российской криминалистической науке данная тематика пока не получила требуемого освещения и развития. О разработке специализированных программных средств соответствующего целевого назначения также не известно.

Поэтому исследования по разработке методов, ориентированных на изучение внешних проявлений речемыслительных процессов в условиях сокрытия человеком информации о событиях прошлого, и создание программных средств для оценки возможности практической реализации этих методов, являются актуальными. Следует учесть также и то, что в проводимых ранее исследованиях под руководством профессора Носенко Э.Л. [4] исследования проводились фактически вручную без применения каких-либо средств автоматизации и программного анализа.

Постановка задачи

Для изучения и оценки возможности разработки метода акустико-психолингвистического анализа устной речи было разработано соответствующее средство программного анализа устной речи (далее – программное средство) по нескольким характеристикам (частота основного тона, динамика распределения фонем, количество и длительность пауз) и проведён ряд экспериментов.

1. Описание разработанного средства программного анализа

Работа с программным средством построена следующим образом. Пользователь загружает фонограмму речи в формате *.wav. в программное средство, и программа автоматически осуществляет следующие действия:

- делит фонограмму анализируемой речи на равные фрагменты (величина которых устанавливается оператором),
- подсчитывает количество фонем в каждом анализируемом фрагменте,
- показывает динамику частоты основного тона фонограммы,
- выделяет в каждом фрагменте и подсчитывает паузы различной длительности, разделенные на шесть групп (Рис. 1).

 - от 200 мс до 400 мс	 - от 800 мс до 1000 мс
 - от 400 мс до 600 мс	 - от 1000 мс до 1200 мс
 - от 600 мс до 800 мс	 - более 1200 мс

Рис. 1. Классификация пауз по длительности.

Для корректной работы программного средства анализа устной речи требуется предварительная очистка фонограммы от шумов. Предварительная очистка частично выполняется средствами аудиоредактора Sony Sound Forge автоматически в процессе записи, а также встроенным средством Noise Reduction Plugin со стандартными настройками. В некоторых случаях проводится дополнительная шумочистка в ручном режиме с использованием специализированных инструментов аудиоредактора Sony Sound Forge. Условия звукозаписи в эксперименте близки к студийным.

Программное средство анализа устной речи написано на языке программирования C++ в среде разработки Embarcadero C++ Builder XE3. Интерфейс программного средства позволяет визуально наблюдать результат анализа исследуемой фонограммы (Рис. 2).

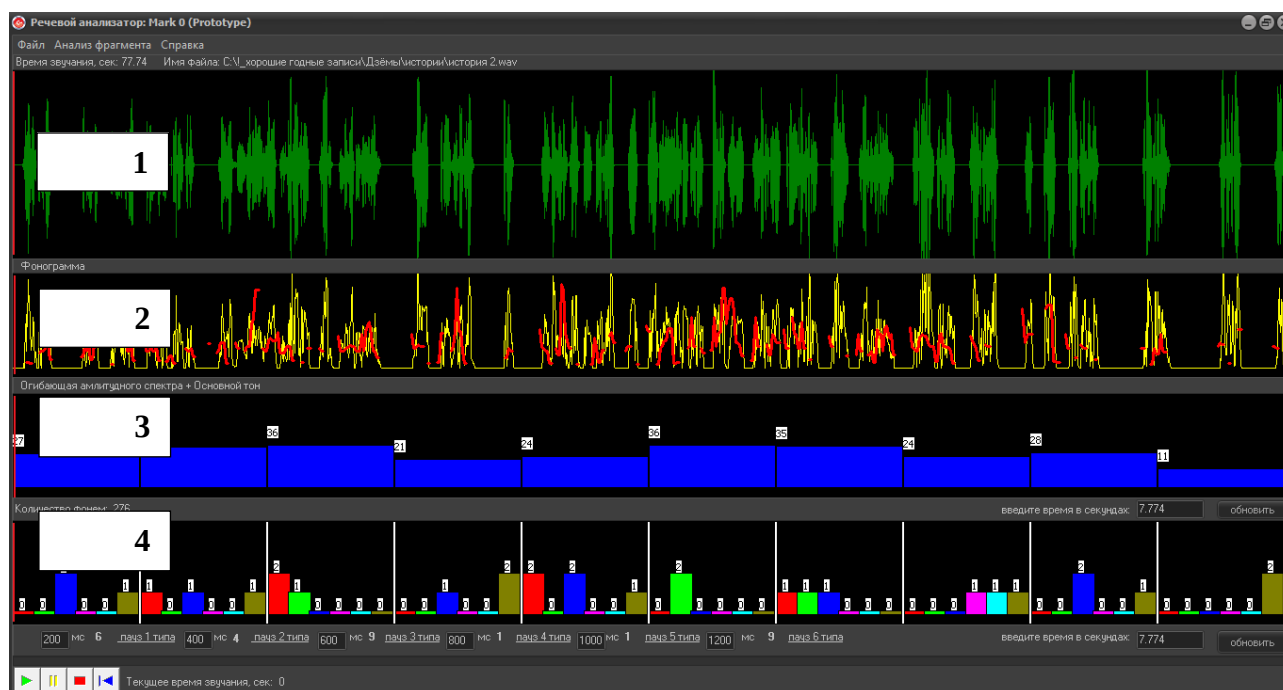


Рис. 2. Интерфейс программного средства.

На участке экрана, отмеченной цифрой «1», представлена исследуемая фонограмма речи. По оси абсцисс отложено время, по оси ординат амплитуда в децибелах. На участке, отмеченном цифрой «2», представлены амплитудная огибающая фонограммы (желтый цвет) и график изменения частоты основного тона (красный цвет). По оси абсцисс также отложено время; по оси ординат амплитуда в децибелах для огибающей спектра и частота в герцах для графика частоты основного тона. На участке «3» – представлено распределение количества фонов на каждом из равных по длительности фрагментов речи. По оси абсцисс отложено время, по оси ординат представлено количество фонов на участках равной длительности. На участке «4» – представлена гистограмма пауз шести указанных выше (Рис. 1) групп на участках равной длительности. В программное средство встроен медиаплеер, который позволяет воспроизводить записанную фонограмму с любого её участка.

2. Экспериментальные исследования

Для проверки работоспособности программного средства в части корректного определения частоты основного тона было осуществлено сравнение динамики основного тона одной из фонограмм, установленной с помощью программного средства (Рис. 3), с результатами спектрального частотного анализа той же фонограммы, выполненного профессиональным аудиоредактором Adobe Audition CS5.5 (Рис. 4).

Фонограмма представляла собой запись нескольких слов, произнесенных одним диктором с различной тональностью голоса. Сравнив Рис. 3 и 4, нетрудно заметить, что динамика частоты основного тона, выполненная различными программными средствами, была определена, фактически, идентично.

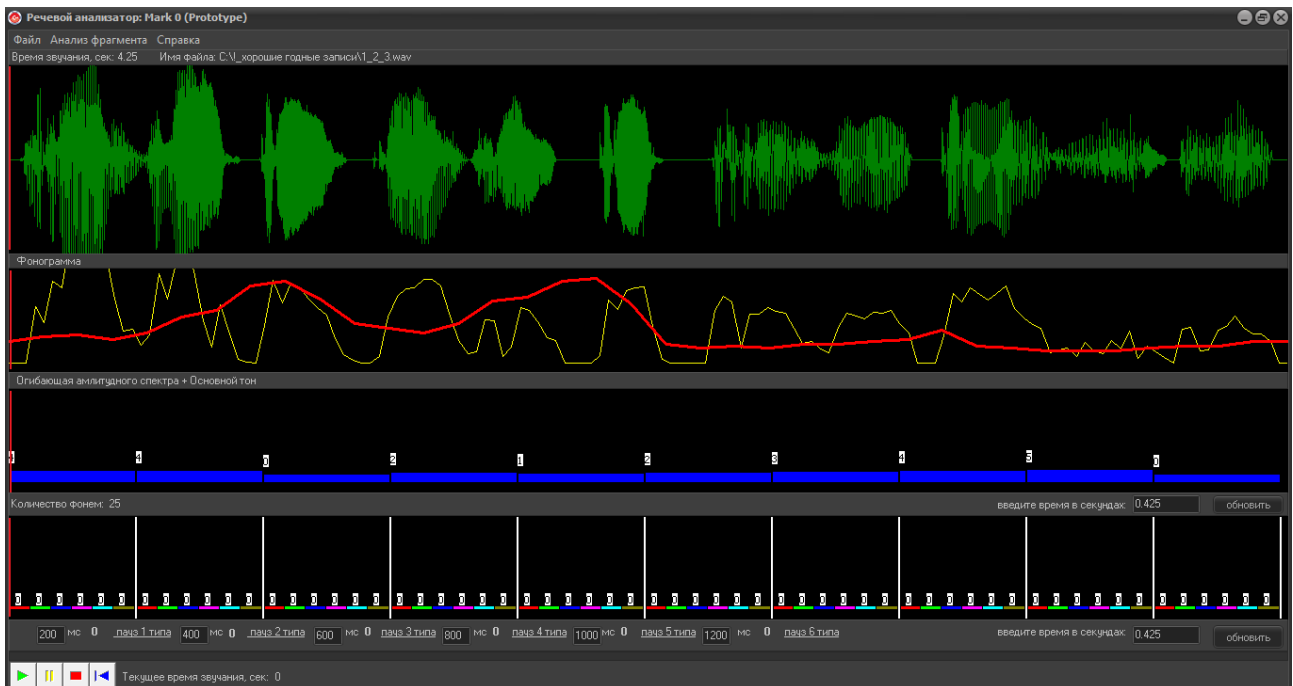


Рис. 3. Оценка динамики частоты основного тона с помощью программного средства анализа устной речи.

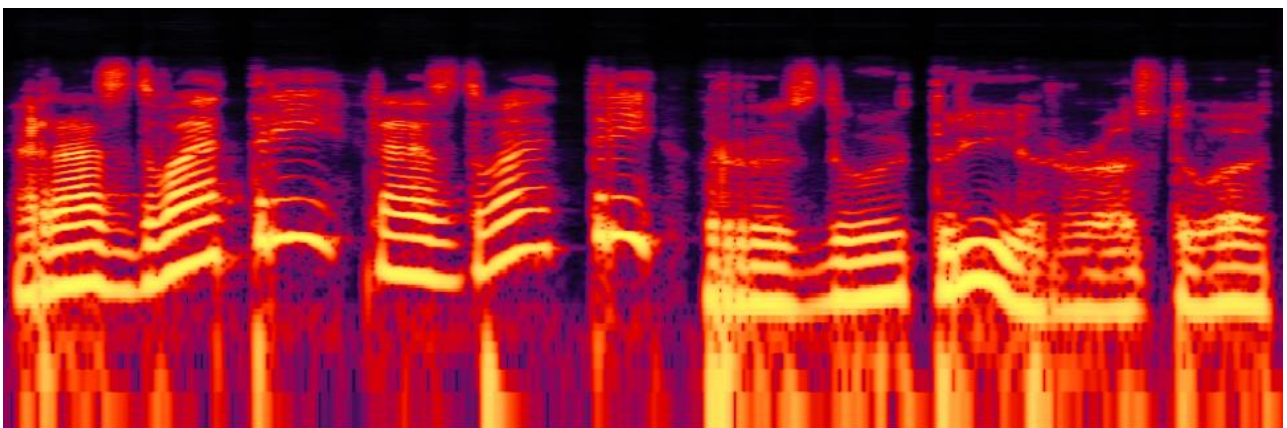


Рис. 4. Оценка динамики частоты основного тона с помощью аудиоредактора Adobe Audition CS5.5.

Для изучения различий в устной речи, реализуемой в различных условиях, были исследованы фонограммы одного и того же текста, который зачитывался диктором дважды – в нормальном (спокойном) темпе и в максимально возможном для него.

При быстром чтении динамика распределения пауз заметно смещается в область пауз меньшей длительности, и количество пауз также уменьшается. При этом сравнительно длинные паузы (Рис. 5) с увеличением темпа речи становятся значительно короче, «уходят» из интервала более 200 мс, либо вовсе исчезают (рис. 6).

В программном средстве реализована функция анализа выделенного фрагмента фонограммы: в рамках «окна» анализа выполняется расчёт длительности фрагмента в миллисекундах, подсчёт количества фонем и скорости речи в данном фрагменте, а также проводится классификация пауз по длительности с их подсчетом.

Проведенные эксперименты наглядно продемонстрировали различие в реализации речи во время чтения и в ходе произвольного монолога на незначимую для диктора тему. На Рис. 7 видно, что при чтении в основном преобладают короткие паузы длительностью от 200 до 600 мс, а при произвольном монологе (Рис. 8) начинают превалировать длительные паузы от 600 до 800 мс, и более 1200 мс; при этом возрастает общее число пауз. Видно, что на выделенных фрагментах существенно различаются скорости речи: при чтении – 5,67 фонем/сек, а при произвольном монологе – 3,91 фонем/сек.

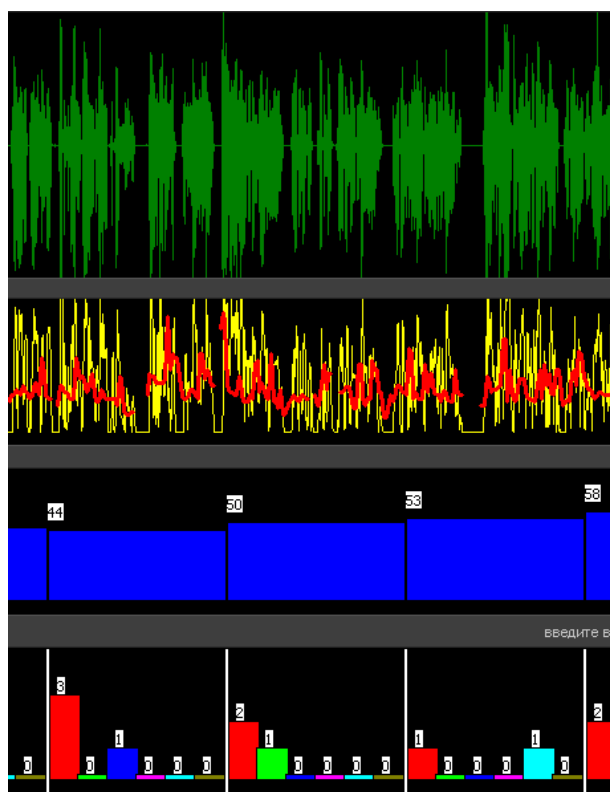


Рис. 5. Чтение в спокойном темпе.

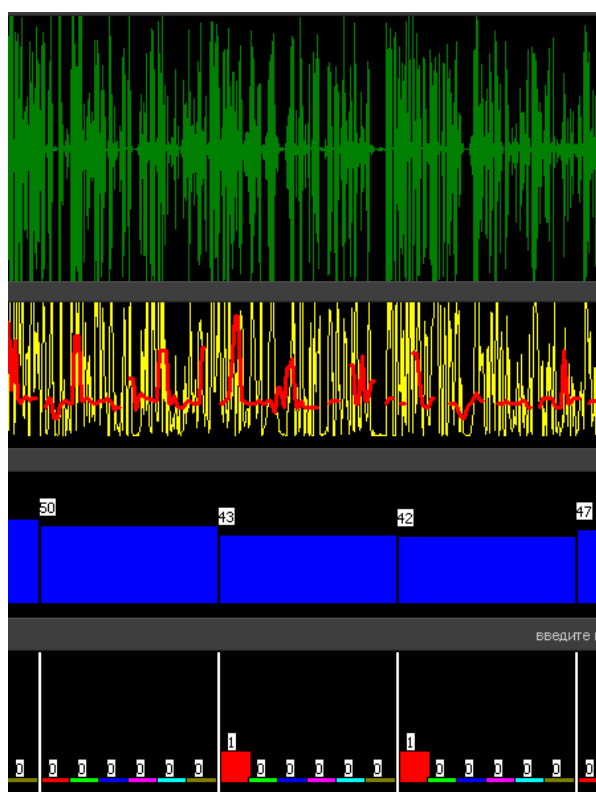


Рис. 6. Чтение в максимально возможном темпе.

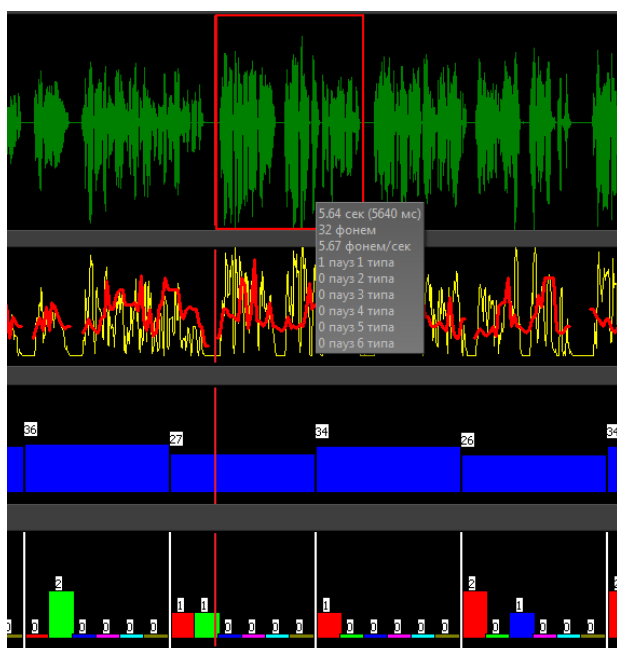


Рис. 7. Чтение в спокойном темпе.

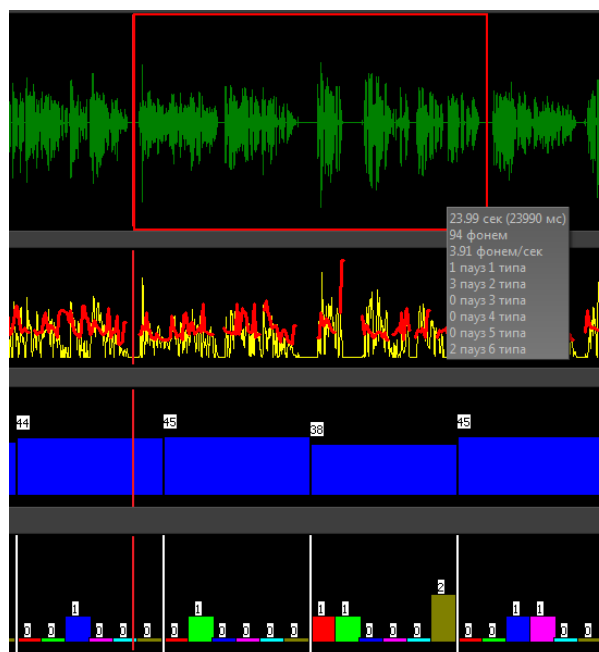


Рис. 8. Произвольный монолог на незначимую для диктора тему.

Далее, эксперименты показали отличия в реализации подготовленной речи (пересказ только что прочитанного текста) и подготовленной речи с включением в неё наспех вымышленного фрагмента неподготовленной речи. Обнаружилось, что в пересказе (Рис. 9) преобладают короткие паузы 1, 2 и 3 типов (200 – 800 мс), а длинные паузы (4, 5 и 6 типов) представлены слабо (более 800 мс – 6 пауз, более 1200 мс – 2 паузы). Напротив, в изложении рассказа с вымышленным фрагментом (Рис. 10) распределение пауз иное: 13 пауз 4-5 типов, 5 пауз 6 типа, а относительно коротких пауз (1, 2 и 3 типы) остается практически неизменным.

В процессе пересказа, который длился 65,54 сек., было произнесено 395 фонем; скорость речи составляла при этом 6,02 фонем/сек. Сообщения с вымышленным фрагментом длилось 65,03 сек. и содержало 274 фонемы; скорость речи – 4,21 фонем/сек. Полученный результат наглядно показывает наличие существенного различия в средней скорости речи, порождаемой воспоминанием и вымыслом. Динамика распределения фонем показана в третьей области («синие столбики») рисунков 9 и 10.

Иными словами, показано, что в речи диктора, испытывающего психическое напряжение в связи с необходимостью реализации краткого вымышленного сообщения, проявляется тенденция к заметному возрастанию суммарной длительности пауз при незначительном увеличении их количества. Отражение указанной тенденции в изменении распределения пауз графически показано в четвёртой области Рис. 9 и 10.

Точность подсчета фонем, осуществляемого программным средством, была проверена экспертным путем пересчета фонем в записанном вручную тексте речи. Было установлено, что суммарная погрешность подсчёта фонем составляет от 0,8% до 9,3%, зависит от скорости произношения диктора и, по-видимому, имеет устойчивую величину для каждо-

го диктора. Паузы программное средство определяет и классифицирует по длительности безошибочно.

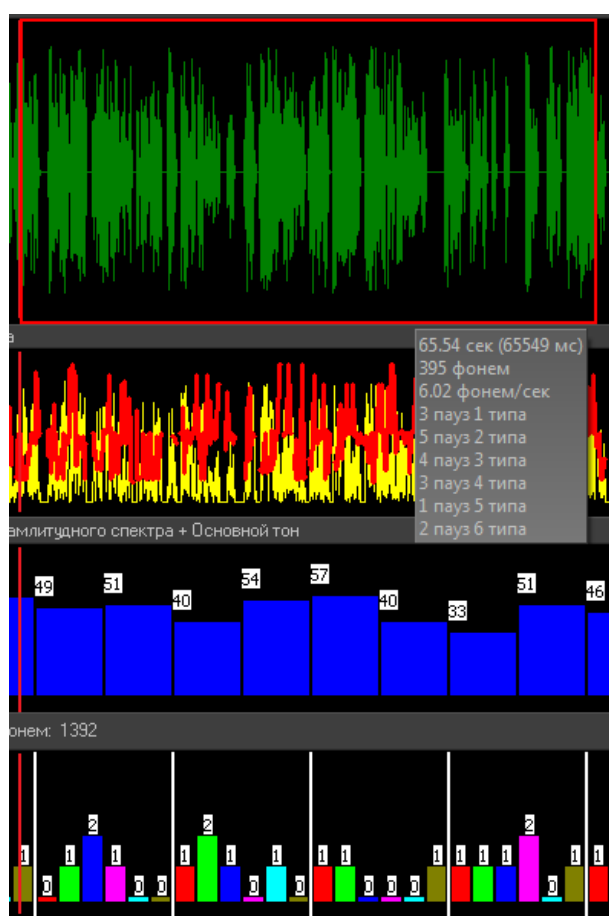


Рис. 9. Пересказ прочитанного.

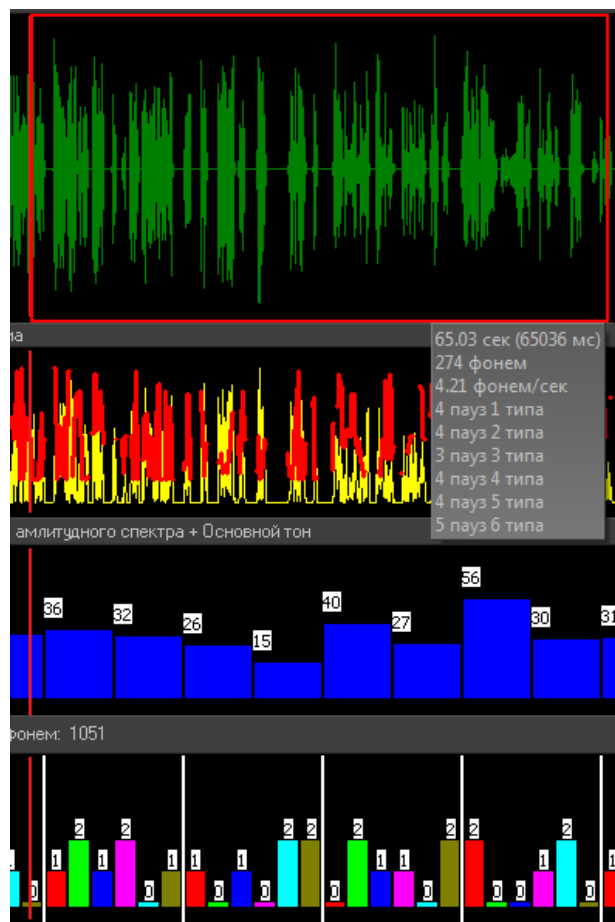


Рис. 10. Подготовленная речь с вымышленным фрагментом.

Заключение

Созданный инструмент программного анализа и проведённые эксперименты показали потенциальную возможность проведения системных исследований по разработке метода диагностики наличия у человека возможно скрываемой им информации путем программного анализа акустико-психолингвистических характеристик его речи.

Список литературы

1. Ларионцева Е.А., Иванова Е.Ю., Стельмашук Н.Н. Голосовые детекторы лжи // Молодежный научно-технический вестник. Электронный журнал МГТУ им. Н.Э. Баумана. 2012. № 6. С. 17.
2. Matsumoto S. A compendium of voice stress research // Polygraph. 2001. Vol. 30. № 2. P. 122-144.

3. Леонтьев А.А., Шахнарович А.М., Батов В.И. Речь в криминалистике и судебной психологии. М.: Наука. 1977. 62 с. С.38.
4. Носенко Э.Л., Холодный Ю.И. Синтаксическое структурирование как неосознаваемый компонент высказывания // Материалы XVIII всесоюзного симпозиума по психолингвистике и теории коммуникации (Тезисы докладов). М.: Институт языкознания АН СССР. 1985. С. 36-37.
5. Салтевский М.В. Криміналістика: Підручник: У 2 ч. Харків: Основа. Ч. 1. 1999. 416 с. С. 261-273.

Evaluating the Possibility of Oral Speech Software Analysis Application for Research in Developing a Method for Diagnosing the Presence in an Individual's Memory of Concealed by Him Information

06, June 2015

DOI: 10.7463/0615.0780454

Yu. I. Kholodniy¹, S.D. Panin¹, V.A. Belykh^{2,*}

УДК: 159.96

¹Bauman MSTU, Moscow, Russia

²Central Scientific Research Institute of Radio Engineering
n. a. Academician A.I. Berg

^{*}v.a.belykh@gmail.com

The problem of creating oral speech software analysis tools for studying the possibility of developing methods aimed at research of verbal and cognitive processes externalization in the context of an individual's concealing information about past events has been considered. A corresponding tool for oral speech software analysis by several characteristics (fundamental tone frequency, phonemes distribution dynamics, number and duration of pauses) has been developed and described. A series of experiments clearly demonstrating the working results of the created software analysis tool and showing practical significance of the analyzed oral speech characteristics has been conducted and described. The developed software analysis tool and the obtained results make it possible to proceed with system research in development and practical implementation of the method for diagnosing the presence in an individual's memory of a possibly concealed by him information by means of analyzing the acoustic and psycholinguistic characteristics of his speech.

Publications with words: [acoustics](#), [psycholinguistics](#), [oral speech analysis](#)

References

1. Lariontseva E.A., Ivanova E.Yu., Stel'mashuk N.N. Voice lie detectors. *Molodezhnyi nauchno-tekhnicheskii vestnik MGTU im. N.E. Baumana* = *Youth Science and Technology*

Herald of the Bauman MSTU, 2012, no. 6. Available at:

<http://sntbul.bmstu.ru/doc/490858.html> , accessed 01.05.2015. (in Russian).

2. Matsumoto S. A compendium of voice stress research. *Polygraph*, 2001, vol. 30, no. 2, pp. 122-144.
3. Leont'ev A.A., Shakhnarovich A.M., Batov V.I. *Rech' v kriminalistike i sudebnoi psikhologii* [Speech recognition in criminology and forensic psychology]. Moscow, Nauka Publ., 1977, p. 38. (in Russian).
4. Nosenko E.L., Kholodnyi Yu.I. Syntactic structuring as the unconscious component of statements. *18 vsesoyuznyi simpozium po psikholingvistike i teorii kommunikatsii: tez. dokl.* [Abstracts of 18th all-Union Symposium on Psycholinguistics and Communication Theory]. Moscow, Institute of Linguistics of USSR Academy of Sciences Publ., 1985, pp. 36-37. (in Russian).
5. Salteviskii M.V. *Kriminalistika. U 2 ch. Ch. 1* [Forensics. In 2 pts. Pt. 1]. Kharkiv, Osnova Publ., 1999, pp. 261-273. (in Ukrainian).