

Идентификация дикторов по цифровым фонограммам

Галяшина Е.И.

Московская государственная юридическая академия имени О.Е. Кутафина.
Россия, 123995 Москва, Садовая-Кудринская, дом 9.
Тел. (499)244-85-24. Факс: (499) 244-87-76. E-mail: galyashina@rambler.ru

Рассмотрены проблемы, возникающие в практике судебно-экспертной идентификации диктора по цифровым фонограммам. Автор рассматривает особенности цифровых записей речевых сигналов, которые фигурируют по различным категориям уголовных дел в качестве доказательств. Основное внимание уделено влиянию искажений оцифрованного речевого сигнала на достоверность идентификации, а также проблеме сопоставимости фонограмм, записанных на различных цифровых устройствах. Делается вывод о необходимости разработки новых экспертных методик, учитывающих специфику цифровой аудиозаписи.

Техническая революция привела к тому, что аналоговая аудио- и видеозапись повсеместно уступает место цифровой записи, для ведения переговоров используются цифровые каналы связи, мобильные телефоны, интернет-телефония, спутниковая связь. Правоохранительные органы постепенно переходят на цифровую звуко- и видеозапись для фиксации следственных действий и при проведении оперативно-розыскных мероприятий. Цифровые фонограммы записываются при помощи цифровых диктофонов, сетевых телефонов, многоканальных цифровых регистраторов, цифровых видеокамер, записывающих устройств на DVD и мини-дисках. На разнообразных носителях цифровые фонограммы приобщаются к уголовным и гражданским делам в качестве вещественных доказательств. Удобство устройств цифровой звукозаписи очевидно — малые размеры устройств, возможность записи длительных переговоров, приемлемое (по мнению потребителей) качество получаемых звукозаписей, низкие требования к пользователю таких устройств.

Казалось бы, преимущества цифровой звукозаписи неоспоримы. Однако экспертная практика показывает, что кроме организационно-технических и процессуальных сложностей приобщения носителей цифровых записей к материалам дела, проблемы возникают и при проведении экспертных исследований аудиофайлов речевых сигналов. Уголовное и гражданское судопроизводство требует проверки подлинности и достоверности фонограмм, выступающих в качестве вещественных доказательств. Основным процессуальным способом такой проверки является судебная фоноскопическая (или фонографическая) экспертиза. Уже более 40 лет фоноскопические экспертизы проводятся во всех экспертных учреждениях правоохранительных органов и в ряде негосударственных экспертных учреждений. При проведении таких экспертиз следствие и суд интересуют вопросы установления подлинности фонограммы и идентификации конкретного диктора по фонограммам устной речи [1].

Анализ экспертных заключений последних лет показывает, что наибольшие трудности возникают при идентификации дикторов по цифровым фонограммам. При этом эксперты не учитывают специфики цифровых фонограмм, цифровой связи, самого процесса

цифровой звукозаписи, особенностей цифровых устройств звукозаписи и её носителей.

В настоящее время в правоохранительных органах для идентификации дикторов используются две автоматизированные системы идентификации дикторов — «Диалект» и «Фонэкси». Федеральным межведомственным координационно-методическим советом по проблемам экспертных исследований паспортизована только одна методика идентификация лиц по фонограммам русской речи на автоматизированной системе «Диалект», 1996 г. [2]. При проведении экспертного исследования идентификация диктора производится на основе измерения таких параметров, как частота основного тона, частоты формант, длительность сигналов и вычисления их производных. Всего на основе результатов акустических измерений вычисляется около 600 акустических признаков голоса и речи. Полученные результаты на спорных фонограммах сравниваются с образцами голоса и речи подозреваемого лица.

Важно подчеркнуть, что названная методика разрабатывалась применительно к исследованию аналоговых фонограмм, записанных в основном по телефонному тракту, или непосредственно на магнитную ленту при помощи аналоговых магнитофонов. Сейчас на экспертное исследование всё больше поступает фонограмм, полученных на цифровых диктофонах, записей телефонных переговоров по сотовой мобильной связи и цифровых регистраторов. В качестве образцов сравнения при таких исследованиях выступают аналоговые фонограммы допросов подозреваемых, цифровые записи с диктофонов, DVD-дисков и видеомагнитофонов. При этом новая редакция методики [3], обозначив наличие проблемы искажения речевых сигналов при их цифровой обработке [4], трудности исследования цифровых фонограмм не разрешила.

Цифровые фонограммы сегодня вызывают у судебных экспертов существенные затруднения не только в выборе оптимальных методов и методик их исследования, но и криминалистической оценке полученных результатов, их квалификации в соответствии с установленными терминологическими стандартами и экспертными методиками. Это связано с тем, что достижения в цифровой обработке и машинном синтезе речи в принципе позволяют осуществить фальсификацию не только содержания, но и всех других компонентов речевого сигнала, поступающего на вход различных систем принятия решений или контроля доступа и фиксируемого на цифровой носитель в целях доказывания. Проблема усугубляется тем, что современные системы цифровой регистрации, обработки звуковых сигналов, специализированные процессоры линейного, нелинейного монтажа, в том числе компьютерного, позволяют осуществлять многообразные манипуляции с речевым сигналом, привнося в его форму и содержание существенные изменения. Учитывая тот факт, что за фальсификацию доказательств установлена уголовная ответственность, вопрос о выявлении таких признаков далеко не праздный. С юридической точки зрения фальсификация — это искусственное создание доказательств, свидетельствующих в пользу обвиняемого или потерпевшего, истца или ответчика. Применительно к задачам исследования цифровых фонограмм, это искусственно составленная фонограмма, содержащая искажённую информацию о документируемом звуковом событии, включая искажение речевого сигнала, приводящее к сходству до степени смешения идентификационных признаков разных

дикторов, либо смещения индивидуализирующих свойств естественной и синтезированной речи.

Современные цифровые звукозаписывающие и передающие устройства конструктивно построены так, что при оцифровке речи они убирают из акустического сигнала значительное количество криминалистически значимой информации об индивидуальных свойствах голоса и речи диктора, условиях формирования акустического сигнала. В некоторых сотовых телефонах вместо речи человека передаётся набор кодовых сигналов, являющихся не нативной (native) речью, а ее математической моделью. При этом от типа используемого алгоритма кодирования речи и его реализации в конкретной модели устройства зависит «похожесть» оригинальной речи говорящего и того, что получает абонент на приёмном устройстве. Таким образом, на выходе создается искусственным образом синтезированный сигнал по некой математической модели, что затрудняет опознаваемость голоса на слух в условиях помех. Похожие голоса, одной группы, не имеющие ярких различительных особенностей, могут кодироваться примерно одинаково, а потому окажутся сходными до степени смещения при их слуховой перцепции экспертом. Лингвистические признаки, выявляемые на слух, зависят от многих факторов, включая индивидуальные качества эксперта, чувствительность его слухового аппарата и опыт. Совершенно очевидно, что когда при цифровой обработке речи появляются дополнительные шумы квантования и иные специфические искажения, достоверность выделения на слух тонких фонетических особенностей оказывается весьма сомнительной.

Именно в такой особенности передачи речи по сотовой связи и лежат предпосылки совершения так называемого «телефонного мошенничества», когда при общении по сотовой связи вероятность ошибочного опознания «чужого» голоса как «своего» достаточна высока.

Дополнительные сложности возникают при экспертном исследовании аналоговых фонограмм, полученных путем перезаписи с исходных цифровых фонограмм, в том числе с привнесенными изменениями. В итоге получается аналоговая фонограмма с цифровым монтажом, микшированным, синтезированным или компилированным речевым сигналом. При оцифровке и обработке такой фонограммы появляются дополнительные искажения, влекущие ошибку результатов сравнения идентификационных признаков.

При использовании для звукозаписи бытовых цифровых диктофонов часто в расчет принимаются только их малые размеры и вес. Не учитывается, что при такой звукозаписи цифровые диктофоны конструктивно устроены так, что все низкочастотные сигналы ниже 300–500 Гц просто отсекаются. На практике можно встретить экспертные заключения, где эксперт «измерял» для таких фонограмм частоту основного тона голоса и низкочастотную составляющую 50 Гц. О надёжности идентификации диктора в этом случае говорить не приходится.

Другая проблема. Портативные цифровые диктофоны и сотовые телефоны для записи длительных телефонных переговоров, как правило, используют различные алгоритмы сжатия (кодирования) сигналов. Чаще всего это так называемое сжатие с потерей качества. На основе психоакустической модели из оцифрованного сигнала удаляется большое количество важнейшей криминалистически значимой информации об объекте исследования (например, кодирование MP3). Аналогичная ситуация с цифровой видеозаписью. Там идет кодирование по различным алгоритмам MPEG.

Ещё более серьёзная ситуация с мобильной связью. В цифровой фонограмме, полученной по мобильной связи идентификационному исследованию, по сути, подлежит не нативный, то есть исходный, естественный сигнал, естественная человеческая речь, со

всеми индивидуальными свойствами присущими конкретному диктору, а искусственно синтезированный сигнал, из которого алгоритмами кодирования (или сжатия, цифровой обработкой, например, ACELP) физически исключены идентифицирующие диктора особенности. При данном способе кодирования по сотовой связи передается не сам сигнал, а его математическая модель — страница в кодовой книге, и эксперты, очевидно, не обладая знаниями специфики сотовой связи, часто механически сравнивают синтезированную (машинную) речь с речью живого человека. Говорить о качестве такого сравнения также не приходится.

Сравнение сопоставимых по своим параметрам речевых сигналов на спорной фонограмме и сравнительных образцах нарушает основной методологический принцип криминалистической и экспертной идентификации — принципа сопоставимости исследуемых объектов в системе признаков, по которым осуществляется сравнение и идентификация диктора.

Другая особенность сотовой связи. Обработка речи в стандарте GSM [5] осуществляется в рамках принятой системы прерывистой передачи речи DTX (Discontinuous Transmission), которая обеспечивает включение передатчика только тогда, когда пользователь начинает говорить и отключает его в паузах и в конце разговора. Система DTX управляет детектором активности речи VAD (Voice Activity Detector), который обеспечивает обнаружение и выделение интервалов речи с шумом и без шума речи даже в тех случаях, когда уровень шума соизмерим с уровнем речи. В состав системы прерывистой передачи речи входит также устройство формирования комфортного шума, который включается в паузах речи, когда передатчик отключён. Экспериментально показано, что отключение фонового шума на выходе приёмника в паузах при отключении передатчика раздражает абонента и снижает разборчивость речи, поэтому применение комфортного шума в паузах считается необходимым. Процесс DTX в приёмнике включает также интерполяцию фрагментов речи, потерянных из-за ошибок в канале.

Все имеющиеся на сегодня детекторы тона имеют выраженный дефект — ложное детектирование тона в интенсивных шумовых сигналах. Вследствие этого шипящие звуки устной речи и многие акустические сигналы шумового характера делаются более звонкими. В алгоритме также используется так называемая «постфильтрация» — заглаживание специальным фильтром всех дефектов восстановленного (синтезируемого) речевого сигнала на выходном конце. Как видно из вышесказанного, идентификация диктора по физическим признакам речи, диагностика акустического окружения и лингвистический анализ фонетических признаков речи в таком сигнале значительно затруднены. Проблема имеет принципиальный методологический характер и заключается в том, что исследованию подлежит синтезированный речевой сигнал, из которого алгоритмом кодирования исключены существенные идентифицирующие диктора, обстановку и канал связи особенности. Без решения вопросов достоверности воспроизводимого сигнала и границ допустимости его искажений специально для сигнала переговоров в GSM канале экспертное исследование не может быть полным.

Кроме того, каждая фонограмма в GSM канале по своей сути является «смонтированной» из цифровых реплик абонентов, между которыми производится вставка искусственного сигнала «комфортного шума». По ГОСТ 13699-91

это подпадает под понятие монтажа фонограммы. Смонтированный таким образом сигнал, включая места стыков — монтажных переходов между передаваемым речевым сигналом и синтезируемым шумом, сглаживается специальным постфильтром. И самое главное — по каналу связи передаётся не сама речь, а некий набор кодов и символов, по которым на оконечном аппарате восстанавливается (синтезируется) сигнал, похожий на исходный, но который не является исходным сигналом, будучи его приближённой моделью. В цифровых регистраторах сигнал подвергается дополнительному сжатию по определённому алгоритму. Результирующий сигнал, подвергшийся кодированию, многократному сжатию и обработке, переписывается на съёмный носитель и поступает на экспертное исследование. При этой экспертизе, как правило, не предоставляется ни оригинал фонограммы, ни устройство звукозаписи, ни информация об условиях передачи сигнала по каналам связи, ни сведения о технологической цепочке записи-перезаписи аудиофайлов.

Следовательно, традиционные экспертные методики исследования фонограмм на предмет идентификации диктора малоприменимы к современной цифровой звукозаписывающей технике и средствам мобильной связи. Экспертное решение о принадлежности голоса и речи, записанного на фонограмме конкретному проверяемому (подозреваемому) лицу, должно основываться на положениях, дающих возможность проверить в условиях судопроизводства обоснованность и достоверность сделанных выводов на базе общепринятых научных и практических данных. Соблюдение названного требования возможно, когда в экспертном заключении с достаточной для воспроизведения подробностью и научной полнотой описан ход исследования и последовательность действий эксперта, дана ссылка на применённую экспертную методику. При этом под методикой следует понимать систему предписаний (категорических или альтернативных) по выбору и применению в определённой последовательности и в определённых или создаваемых условиях методов и средств решения экспертной задачи.

В данной ситуации необходимо проведение комплекса научных исследований в данном направлении и разработка соответствующих экспертных методик, учитывающих особенности цифровых каналов связи и цифровых фонограмм как объектов экспертного исследования. При этом необходимо задействовать как научный потенциал экспертных учреждений, так и академическую науку, как это было в 60–80-е годы при разработке методики «Диалект».

В заключение необходимо отметить, что проблема идентификации дикторов по цифровым фонограммам существует, она пока не решается, но и замалчивать её контрпродуктивно — без надёжной и общепринятой методологии исследования цифровых фонограмм можно вообще потерять цифровые фонограммы как вещественное доказательство. Поэтому необходимо в кратчайшие сроки разработать и внедрить комплекс научно-методических исследований и организационных мероприятий для всестороннего исследования цифровых фонограмм, полученных в разных условиях на разных технических средствах. К такой работе в обязательном порядке кроме ведомственных учёных и экспертов необходимо привлечь академическую науку, учёных-речевиков, юристов и законодателей.

ЛИТЕРАТУРА

1. Галяшина Е.И. Актуальные проблемы экспертизы цифровых фонограмм. /Теорія та практика судової експертизи і криміЗборник наукових трудов. Випуск 8. Харків: «право», 2008. С. 248–257.

2. Попов Н.Ф., Линьков А.Н., Кураченкова Н.Б., Байчаров Н.В. Идентификация лиц по фонограммам русской речи на автоматизированной системе «Диалект». М.: Войсковая часть 34435, 1996.
3. Кураченкова Н.Б., Байчаров Н.В., Ермакова М.А. /Под ред. В.М. Богданова. Идентификация лиц по устной речи на русском языке. Методика «Диалект». Пособие для экспертов. Издание 2-е, переработанное и дополненное. М., 2007.
4. Голощапова Т.И., Захаров А.Г., Богданов И.Е. О влиянии методов цифрового кодирования на идентификацию диктора по голосу//Криминалистика XXI век: Материалы научно-практической конференции. М.: ГУ ЭКЦ МВД России, 2001.
5. Галяшина Е.И., Галяшин В.Н. Цифровые фонограммы как судебное доказательство, Воронежские криминалистические чтения. Воронеж: Изд-во Воронежского гос. университета. Вып. 8. 2007. С. 71–99.