
Глава 1. КРИМИНАЛИСТИЧЕСКИЕ ФОНОСКОПИЧЕСКИЕ ИССЛЕДОВАНИЯ И ЭКСПЕРТИЗЫ

§1. История фоноскопии и ее методологии

Установить “дату рождения” криминалистической фоноскопии так же невозможно, как и решить подобный вопрос о времени рождения и авторстве любого другого ярко выраженного направления в науке, технике, медицине или в культуре в целом. Часто популяризация и широкое распространение ранее малоизвестных научных данных многими воспринимаются как открытие, причем авторство зачастую приписывается лицу, обнародовавшему эти данные. Но ни одно научное открытие или разработка не происходят в одиночку. Созревание нового в культуре общества длится дольше жизни одного человека. Любой выдающийся ученый, изобретатель или исследователь теснейшим образом связан со своими коллегами, живущими не только в его время, но творившими задолго до него. И неожиданное появление новой идеи, задумки или решения объясняется активностью и разнообразием культурных связей (в широком смысле) между учеными во все времена. В этом плане мир ученых можно сравнить с мозгом человека, в котором постоянно идет обмен информацией между отдельными клетками. И благодаря этому широкому обмену образуется нечто загадочное, называемое нами интеллектом.

Знакомство с публикациями по вопросам истории фоноскопии вообще и криминалистической в частности показывает, что разные авторы по-своему видят корни ее развития. Почему-то чаще всего их связывают с некоторыми значимыми политическими событиями, а не с разработкой научных, технических, методологических и иных основ, фактически дающих возможность реального решения фоноскопических проблем в той или иной общественной ситуации. Наиболее ярким, относительно свежим примером в этом отношении является взгляд на историю криминалистической фоноскопии, изложенный в монографии Х.Холлиена [92]. В ней, пытаясь определить лидерство в новой отрасли - криминалистической фоноскопии, сначала с иронией упоминается история жизни фоноскопической лаборатории, описанная в романе А.И.Солженицына “В круге первом”. При этом автор затрагивает лишь большие политические проблемы и совершенно не останавливается на сути вопросов криминалистической фоноскопии. Затем, демонстрируя большой временной разрыв, автор указывает на то, что следы пер-

вого обсуждения допустимости применения “аурально-перцептивного” (т.е. на слух) свидетельского показания уходят на несколько столетий назад в Англию, где в 1660 г. рассматривалось дело некоего Вильяма Хьюлета. Реально же свидетельские показания об идентичности речи и голоса стали приниматься судами США и Великобритании наравне с другими уликами, начиная с 1907 г.

Неосведомленному читателю может показаться, что идентификация личности по устной речи решалась уже давно, и если учесть современные научно-технические достижения, то никаких сомнений в разрешимости этого вопроса вообще быть не может. Но несмотря на историческую достоверность отмеченных событий, действительно научного решения вопроса не было в обоих упомянутых случаях. Следует признать, что и в настоящее время в проблеме идентификации личности по речи больше вопросов, чем достоверных ответов. Интуитивно большинство специалистов и обывателей чувствуют, что проблема идентификации личности по устной речи имеет положительное решение, но привести бесспорное доказательство в полном объеме и тем более предложить надежную процедуру (методику) идентификации личности по речи еще никому не удалось.

Основная задача криминалистической фоноскопии традиционно рассматривалась как идентификация личности по устной речи. Речь и голос человека так же, как и почерк, настолько индивидуальны, что даже у специалистов возможность такой идентификации обычно не вызывает сомнения. Объясняется это тем, что идентификация личности по голосу и речи повседневно осуществляется каждым. Однако если почерковедение заняло в криминалистике твердое положение, то фоноскопические исследования все еще никак не могут окончательно определиться. Причина здесь, видимо, кроется в том, что фоноскопия по сравнению с идентификацией личности по письменной речи существенно более разноплановая. И поэтому сколько специалистов разных отраслей знаний - столько и мнений. Однако в отличие от письменной речи, которая является объектом исследования двух видов криминалистических экспертиз - почерковедческой и автороведческой, устная речь является, по крайней мере, объектом трех видов исследований - артикуляционного, голосового и автороведческого. Первые два исследования еще не имеют своего четкого названия, что характеризует современный уровень развития криминалистической фоноскопии, которая еще не разделена по предметам исследования, как в письменной речи, и поэтому занимается сразу всем - идентификацией личности и по артикуляции, и по голосу, и по тексту, идентификацией магнитофонограмм, исследованием признаков монтажа или копирования фонограмм и др.

Если следовать логике разграничения понятий, принятой в почерковедении и автороведении, то задачу идентификации личности по устной речи можно разделить на идентификацию по речевому сигналу и по тексту. Исходя из сказанного, правильнее всего было бы определить основную за-

дачу идентификационного исследования личности в фоноскопии как идентификацию личности по речевому сигналу.

В свою очередь, поскольку сам речевой сигнал отражает две относительно самостоятельные группы индивидуальных признаков, то их можно было бы разделить на артикуляционные и голосовые. Поэтому можно говорить о двух задачах: идентификации личности по артикуляции и по голосу.

Если сравнить уровень современного развития почерковедческих криминалистических исследований и фоноскопических, то явное превосходство оказывается у первых. Произошло это, по-видимому, потому, что рукописные тексты и подписи издревле применялись как вещественные доказательства того или иного факта, и общество уже давно приняло эту бумажную форму информационной технологии. А новая безбумажная технология, к каковой относится и магнитная запись, хотя и пробила себе дорогу, но еще находится в стадии закрепления.

Более предметно и объективно об идентификации личности по речевому сигналу начали говорить лишь в середине двадцатого столетия, когда магнитофон стал применяться сначала в военных целях, а затем и во всех сферах информационной деятельности человека.

Первые попытки идентификации по речевому сигналу строились на лингвистическом анализе речи. Они, как правило, были обречены на неудачу, поскольку в юриспруденции очень силен древневосточный принцип: от лжи (уха) - до истины (глаза) - пять пальцев. А поскольку средств объективного анализа речи практически не было и, соответственно, не было методик идентификации, то в последние десятилетия получили распространение толкования уголовно-процессуальных норм, запрещающих применение следов речи на магнитной фонограмме для идентификации личности. Эти запреты коснулись не только судебной практики, но и существенно тормозили все научные исследования в области фоноскопии. Однако несмотря на препоны фоноскопия развивалась и продолжает развиваться, базируясь при этом на надежном научном фундаменте, имеющем длинную историю.

Основы акустической теории речеобразования были заложены Гельмгольцем Германом Людвигом Фердинандом (1821-1894 гг.) - немецким ученым, автором фундаментальных трудов по физике, биофизике, физиологии и психологии. Он впервые раскрыл природу формирования гласных звуков и вывел основные уравнения взаимосвязи геометрии объемных полых фигур и их резонансных частот. Этот первый теоретический шаг в освоении речевой технологии сразу же породил попытку создать механическую говорящую машину. Для этого предлагалась сложная система акустических резонаторов, которая управлялась специальным механизмом примерно так же, как это делает человек. Однако реализовать эту идею до конца не удалось.

Вторым важным практическим шагом в освоении речевой технологии было изобретение французским любителем Шарлем Кро механического

устройства записи звука. Он 30 апреля 1877 г. послал во Французскую Академию наук свое изобретение, названное им “фонограф”, которое позволяло механически записывать звук на вращающийся стеклянный диск, покрытый сажей, а затем фотоспособом переносить звуковые дорожки на светочувствительную хромовую пластинку. В декабре 1877 г. американский изобретатель Томас Эдисон продемонстрировал свой фонограф в действии, в котором звук записывался на обернутый в оловянную фольгу цилиндр. Через 10 лет, в 1887 г., немецкий изобретатель Эмиль Берлинер запатентовал фонограф, в котором он видоизменил идею Кро - запись звука производится на цинковом диске, покрытом тонким слоем воска, а тиражирование осуществляется химическим травлением.

С появлением сначала механических средств звукозаписи - фонографов, а затем и электромагнитных - магнитофонов исследователи получили возможность многократно и разносторонне исследовать одни и те же речевые сигналы, зафиксированные на фонограмме. Это событие родило волну новых исследований речи, но они, как правило, базировались на аудитивном анализе, что, как уже отмечалось, несет большой отпечаток субъективности.

Ключевым этапом развития фоноскопии вообще (не только криминалистической) стало изобретение прибора, позволившего не только многократно воспроизводить один и тот же речевой сигнал, но и увидеть его преобразованную форму, ярко отражающую артикуляционные и голосовые особенности говорящего. Зарождался этот прибор долго. В течение 20-30-х годов нашего столетия во время известного оживления в научном мире предлагалась масса вариантов визуализации речевых сигналов. После долгих испытаний и отбора лучшего способа к концу второй мировой войны наконец-то был разработан прибор, факт существования которого долго хранился в тайне в силу его большого военного значения. Лишь 9 ноября 1945 г. в еженедельном журнале “Science” в статье “Видимые образы звука” Ральф Поттер [101] впервые приоткрыл завесу секретности над новым инструментом, разработанным сотрудниками фирмы “Bell Telephone Laboratories” [106]. Сначала его называли звуковым спектрографом, но позже за ним закрепилось другое название - сонограф.

Появление сонографа, позволившего увидеть одновременно и артикуляционные, и голосовые особенности говорящего, уже к 1944 г. породило скороспелые проекты полного решения проблемы идентификации личности по речевому сигналу. Позже даже появился долго остававшийся модным термин “voiceprint” - “отпечаток голоса” (по аналогии с термином “fingerprint” - “отпечаток пальца”), который обозначал всего лишь один из видов узкополосных сонограмм речевых сигналов, внешне очень напоминающих отпечатки пальцев.

Такое “удачное” терминологическое определение нового понятия сыграло своеобразную роль в дальнейшем развитии криминалистической

фоноскопии. Поскольку в то время дактилоскопия была вне критики, это послужило как бы подменой прямого доказательства возможности идентификации личности по речевому сигналу. Однако терминологическая неточность сыграла крайне негативную роль в научном решении проблемы, поскольку администраторы от науки считали, что фоноскопическая проблема исчерпана, и поэтому прилагать какие-то усилия в этом направлении нет необходимости. Это, в конечном итоге, привело к тому, что научная обоснованность и убедительность выводов экспертов-фоноскопистов зачастую были, мягко говоря, не на должном уровне, и, как следствие, со временем развился обратный процесс недоверия фоноскопической экспертизе.

Прежде чем фоноскопия превратилась в серьезную науку, ей пришлось пройти еще две стадии развития (научно-теоретическую и инструментально-методологическую), прежде чем удалось обеспечить эксперта минимально необходимой технологией исследования речевых сигналов.

Решающий вклад в это развитие внесло не одно поколение выдающихся ученых: Г.Фант, Дж.Фланаган, А.А.Пирогов, В.Н.Сорокин, И.Максимов, В.П.Морозов, С.А.Гельфанд, Р.Юссон, А.В.Оппенгейм, Р.В.Шафер, Л.Р.Рабинер, Дж.Маркел, А.Грей, Р.Отнес, Л.Эноксон, В.Н.Трунин-Донской, Р.К.Потапова, Г.С.Рамишвили, Ю.И.Прохоров, В.И.Галунов, Н.Г.Загоруйко, Х.Холлиен, Т.К.Винцюк, М.А.Сапожков и многие другие. Это выдающиеся теоретики и практики широкого профиля. Среди ученых, тесно занимавшихся проблемами идентификации личности по речи и внесших огромный вклад в их решение, следует упомянуть: Б.Атала, Дж.Доддингтона, Е.Бунге, В.Д.Сердюкова, М.А.Тушишвили, Г.Б.Чикоидзе, А.Розенберга, В.Л.Шаршунского, Ю.Ф.Жарикова и многих других. Юридические проблемы криминалистической фоноскопии лучше всего исследованы в работах таких ученых, как В.А.Снетков, Э.К.Ребгун, А.А.Леви, Ю.А.Горинов и другие.

После первых попыток, предпринятых во время войны, и провала надежд на простое решение проблемы идентификации личности по речи, внимание ведущих ученых было уделено объяснению природы речеобразования. В этой сфере были развернуты фундаментальные исследования, которые к концу 50-х годов позволили Г.Фанту в тесном сотрудничестве с Акустической лабораторией Массачусетского технологического института создать акустическую теорию речеобразования. Переход от артикуляционного описания речи, принятого в лингвистике, к акустическим методам и создание прочной теоретической базы возродили исследования индивидуальных признаков речи.

В начале 60-х годов Г.С.Рамишвили начал системные исследования индивидуальных признаков речи и голоса, привлекая для этих целей аудиторные, экспертные и все доступные инструментальные методы анализа речевых сигналов. Были установлены акустические параметры речевых сиг-

налов, отражающие те или иные индивидуальные особенности речеобразования, оценены их информативность и надежность определения [49].

В это же время за рубежом Керст предпринял новую попытку разработать методику идентификации личности по сонограммам речевого сигнала. Было заявлено, что его методика позволяет считать проблему идентификации личности по речи полностью решенной. Несмотря на то, что идентификационные исследования по сонограммам носили все же более качественный и субъективный характер, за что не раз и обоснованно подвергались жесткой критике, эта методика все же нашла свое применение в криминалистической практике некоторых стран. Основная причина тому - легкость получения сонограмм при отсутствии иных технологически простых способов исследования речевых сигналов. Однако простота получения сонограмм с "отпечатками голоса" далеко не решила всех проблем идентификации в криминалистике. Поэтому, относясь критично к новому и еще далеко не совершенному виду криминалистических исследований, Международная ассоциация по идентификации (IAI - International Association of Identification) потребовала специальной подготовки экспертов-криминалистов по теоретической и практической части идентификации с обязательной сдачей экзамена для получения удостоверения на право производства фоноскопических экспертиз [82]. В 1975-1977 гг. эта методика подверглась жесткой критике со стороны ведущих ученых и специалистов в области речевой технологии [90, 91].

В эти же годы появляется ряд работ, посвященных поиску индивидуальных признаков речевых сигналов. В работах ведущих ученых (Б.Атал, Г.С.Рамишвили, М.А.Тушишвили, Е.Бунге, А.Розенберг, Дж.Доддингтон и другие) приводились новые данные об отражении индивидуальных особенностей говорящего в его речевом сигнале.

После небольшого затишья в рекламной шумихе о всемогуществе ФБР в сфере речевой технологии в середине 80-х годов снова стали появляться сообщения об их "превосходных" достижениях в области идентификации личности по речи. Так, в авторитетнейшем акустическом научном журнале "The Journal of the Acoustic Society of America" появилась статья [96], в которой сообщалось о блестящих показателях работы ФБР по идентификации личности по речи. В ней указывалось, что ошибки идентификации первого рода (исключение тождества голосов одного и того же лица) составляют 0,53%, а ошибки второго рода (отождествление голосов разных лиц) - 0,31%.

С научной точки зрения, это были бы действительно очень высокие показатели надежности идентификации. Однако ведущим специалистам по речи они показались слишком уж хорошими, после чего в том же журнале [104] ими была опубликована статья о нереальности таких показателей. Критика была настолько убедительна, что авторам из ФБР пришлось поместить в журнале ответную статью [95] с пояснением того, что они в действи-

тельности имели в виду. Оказалось, что это не реальные показатели надежности системы идентификации, а показатели применительно к “очевидным случаям”. В остальных же случаях (которых оказалось большинство) эксперты обычно давали ответы типа: “материал не пригоден для исследования” или “ответить на вопрос не представляется возможным”.

В некоторых критических публикациях указывается на уязвимость методики идентификации личности по “отпечатку голоса” в основном по следующим причинам:

а) в методике не сформулированы четкие критерии сопоставления речевых образов, представляемых на сонограммах;

б) из-за отсутствия привязки исследуемых признаков речевого сигнала, проявляемых на сонограммах, к природе речеобразования слабо аргументированы выводы об идентичности личности;

в) заранее не определено, какие признаки, проявляемые на сонограмме, отражают (или в принципе могут отражать) индивидуальные особенности речевого сигнала.

Эти критические замечания по сути своей не отрицают возможности идентификации личности по сонограммам, но они требуют более строгого научного решения задачи с учетом всех факторов, влияющих на получаемый в реальности образ речевого сигнала. При этом указывается на необходимость учета влияния линейных и нелинейных аппаратных искажений речи в процессе ее записи на магнитные носители, психологической ситуации, вида текста, эмоционального состояния говорящего, акустических условий, шумов и помех и прочих факторов, вносящих вклад в вариативность речевого сигнала.

С внедрением измерительно-вычислительной техники и развитием цифровых методов обработки и анализа сигналов стало возможным проводить уже не только амплитудно-частотно-временной анализ, как это делается с помощью сонограмм, но и с высокой точностью измерять целые комплексы практически всех известных акустических параметров речевого сигнала. Исследования проблем опознавания личности по речи значительно расширились. Цифровые методы обработки сигналов полностью сняли ограничения на перечень анализируемых акустических параметров. Появились возможности оценивать индивидуальные особенности строения и функционирования отдельных звеньев органов речи, основываясь на акустической теории речеобразования. Это сняло многие критические замечания, относящиеся к методике идентификации личности по “отпечаткам голоса”.

Современные возможности цифрового анализа сигналов в сочетании с достижениями в области фоониатрии, речевой акустики и смежных отраслей знаний делают проблему опознавания говорящего принципиально разрешимой. Но, несмотря на это, следует признать, что единой универсальной методики опознавания по речи еще нет. Скорее всего такая универсальная

методика и не будет разрабатываться, поскольку практических задач существует очень много, и каждая из них выдвигает свои требования к характеристикам системы опознавания.

Трудно перечислить все факторы, вызывающие изменение голоса человека в речи. По этим изменениям мы определяем, спрашивает или отвечает собеседник, находится ли он в хорошем настроении или чем-то подавлен, здоров или болен, возбужден или находится в нормальном состоянии и т.п. Все это в той или иной мере находит свое отражение в голосе во время разговора. Но даже при такой широкой вариативности голоса мы все равно узнаем говорящего, не видя его. Более того, можем узнавать его голос, не разбирая не только смысла, но и текста высказывания. Понять, каким образом нам это удастся, - значит найти путь решения криминалистической задачи опознавания личности по голосу. Пока эта задача еще очень далека от своего полного решения, хотя многие индивидуальные признаки речи известны [49, 52, 62, 92] и используются в экспертной практике [31, 49, 64, 74, 91, 92, 107].

С акустической точки зрения, речевой сигнал поддается разбиению на четыре группы участков [67], различающихся по характеру источника звука: пауза, голосовой, шумовой и голосовой в совокупности с шумовым источника звука. Такая классификация не является вполне адекватной и требует введения дополнительных групп: импульсных источников, импульсных с голосовым источником и др. Однако она оказалась удобной для проведения первых фундаментальных акустических исследований и развития акустической теории речеобразования [67].

Для криминалистических исследований более удобна упрощенная классификация участков речевых сигналов, включающая в себя три группы интервалов. К *первой группе* относятся интервалы, на которых отражена работа голосовых связок. Эти участки будем называть тональными участками речевого сигнала. К ним относятся участки гласных и звонких согласных звуков. *Вторая группа* - это участки речевого сигнала, на которых амплитуда сигнала не равна нулю, но голосовые связки не работают. Условно их можно назвать нетональными участками речевого сигнала. К ним относятся, например, шипящие и взрывные звуки. *Третья группа* - это участки звуковых пауз. При кажущейся, на первый взгляд, неинформативности звуковых пауз они, тем не менее, в совокупности с тональными и нетональными участками речевого сигнала, составляющими в целом речевой сигнал определенной фразы, также отражают индивидуальные особенности речи [49].

Такая классификация участков речевого сигнала не является общепринятой, однако она оправдана для криминалистических целей. Из всех акустических параметров наибольший интерес для криминалистики представляют следующие параметры:

- наиболее отражающие индивидуальные особенности речи;
- надежно выделяемые из речевого сигнала;

- наименее искажаемые аппаратурой звукопередачи и записи;
- поддающиеся измерению в шумах с высокой степенью точности.

Среди этих параметров можно выделить: периодичность работы голосовых связок, периодичность следования тональных участков речевого сигнала, длительность периодов смычных и смычно-проходных звуков и т.п. Одной из информативных совокупностей таких признаков являются параметры узкополосных спектров тональных участков, отражающих периодичность работы голосовых связок. Высокая информативность этой совокупности обусловлена проявлением в ней индивидуальных особенностей одновременно двух систем - анатомического строения [67, 49] и нейрофизиологического регулирования голосообразования [77].

Звуковые паузы и нетональные участки речевых сигналов менее информативны для решения криминалистических задач. Обычно на практике используются следующие основные параметры этих участков: их длительность и, если проявляются, формантные особенности шипящих звуков в области частот ниже 4-6 кГц. Остальные параметры, например форма и спектр импульса взрывного звука, обычно не используются из-за сильного искажения и принципиальной невозможности допустимо точного оценивания в шумах (в силу их одиночности в окрестности квазипостоянства характеристик искажающих факторов).

Следует отметить, что, имея приемлемую технологию выделения тональных и нетональных участков речевого сигнала, уже можно исследовать индивидуальные особенности их чередования. Однако эти особенности тесно связаны с семантической стороной речевого сигнала и в настоящей работе не рассматриваются. Основное внимание будет уделено исследованию фониатрических особенностей голоса в речи.

В природе речеобразования выделяются три основных фактора, определяющих конкретный вид тональных участков [67, 69, 33, 49]:

- форма импульса голосового возбуждения речевого тракта;
- периодичность следования голосовых импульсов;
- форма речевого тракта.

Подчеркнем, что индивидуальность речевого сигнала в значительной мере проявляется не в статике, а в динамике этих факторов. Указанные факторы являются базисными, определяющими все остальные воспринимаемые слухом человека и аппаратурно измеряемые признаки тональных участков. Например, такой признак, как чистота звучания голоса, обуславливается одновременно стабильностью формы голосовых импульсов и стабильностью периода их следования друг за другом. Тембр голоса определяется формой голосового импульса и формой речевого тракта.

Для криминалистических целей обязательно знание именно базисных признаков, например, формы голосового импульса или стереометрического описания динамики формы речевого тракта. Главное - знать, в каких группах измеряемых акустических параметров базисные группы признаков

находят свое отражение, чтобы криминалистическое исследование голоса в речи было предельно полным. При этом следует различать, какие признаки отражают артикуляционные, а какие - голосовые. Поэтому, опираясь на базисные факторы, определяющие конкретные измеримые характеристики тональных участков, и учитывая то, что в них отражаются те или иные индивидуальные особенности говорящего, можно попытаться разбить акустические признаки тональных участков на три класса, по типу упомянутых выше факторов.

Получающиеся классы признаков не будут пересекаться. Например, такой признак, как смещение измеряемых обертонов от соответствующих гармоник основного тона, зависит одновременно и от формы импульса голосового возбуждения, и от формы речевого тракта, и от частоты основного тона голоса, и, более того, от способа измерения частоты обертонов. Поэтому некоторые измеримые акустические признаки тональных участков могут быть отнесены не только к одному из вышеуказанных классов. На рис.1 приведена общая схема отношения тех или иных практически используемых и перспективных измеримых акустических параметров тональных участков, которые необходимы для решения криминалистических задач.

Как уже отмечалось, измерение и исследование группы признаков, связанных в большей мере с анализом широкополосных спектров тональных участков, развиты достаточно хорошо. Поэтому системы, основанные на анализе широкополосных спектров тональных участков, разработанные, действующие и описанные в работах [49, 3, 18, 55], должны быть включены практически в неизменном виде в разрабатываемую систему криминалистической идентификации личности по речи. Эти системы необходимо лишь адаптировать к используемой измерительно-вычислительной технике таким образом, чтобы при исследовании одного и того же речевого сигнала применять одинаковые формы параметрического описания речевых сигналов.

Из группы акустических параметров тональных участков, отражающих периодичность и постоянство формы голосовых импульсов, в настоящее время исследованы некоторые статистические признаки периодичности следования голосовых импульсов [49, 65]. Результаты исследования частотной модуляции голоса, чистоты звучания голоса, статистических параметров производной контура частоты основного тона и некоторые другие только обозначены в научной литературе [49, 89, 109]. Проведение этих и многих других исследований, изучающих особенности периодичности работы голосовых связок, сильно ограничены возможностями точного измерения соответствующих параметров, в особенности частоты основного тона.

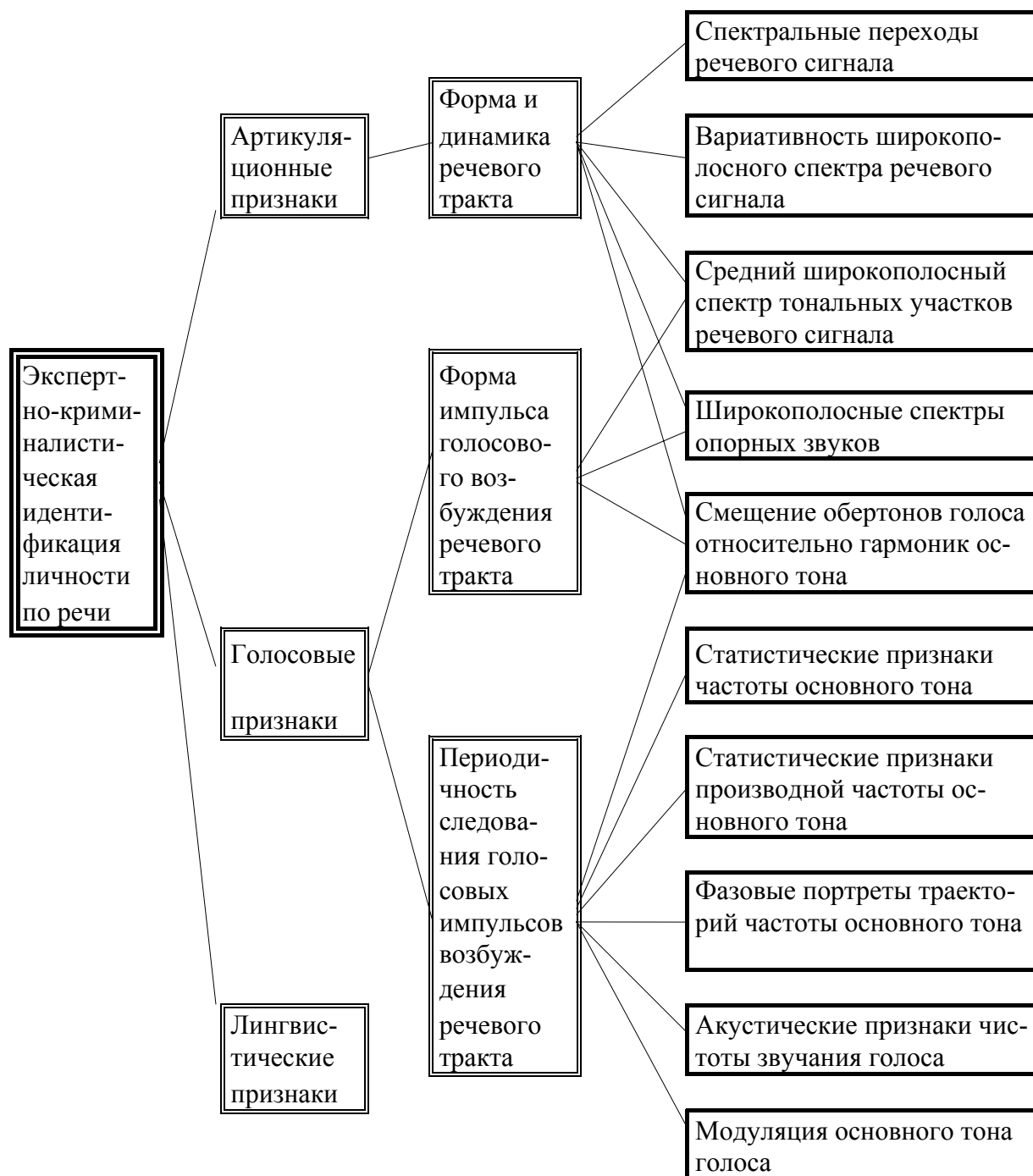


Рис. 1. Отражение в акустических параметрах тональных участков речевых сигналов индивидуальных особенностей речеобразования

Касаясь вопроса автоматизации решения криминалистических задач опознавания личности по речи, отметим следующее. Предусмотреть в алгоритмической схеме исследования все нюансы, начиная от чисто технических особенностей качества фонограммы речи и кончая умышленным искажением голоса говорящим, в настоящее время практически невозможно. Поэтому на данном этапе развития системы криминалистической идентификации личности по речи целесообразно основное внимание уделять из-

мерению и наглядному отражению всех возможных акустических параметров речевых сигналов, которые могут понадобиться в ходе криминалистического исследования. Причем измерения, статистическая обработка и отображение результатов должны производиться с высокой степенью точности, надежности и автоматизации.

§2. Задачи фоноскопии

Как уже отмечалось, криминалистическая идентификация личности по речи включает в себя несколько задач разного типа. Первая задача, возникшая в технических системах и получившая название “верификации дикторов”, формулируется так: действительно ли речевой сигнал, предъявляемый системе распознавания, принадлежит данному лицу или нет. Такие системы, как уже говорилось, обычно используются в специальных контрольно-пропускных пунктах, однако их применение этим не ограничивается. В органах внутренних дел такие системы могут быть полезны для решения вопросов охраны помещений (точнее, снятия с охраны и постановки на охрану объекта по телефону), соблюдения режима на контрольно-пропускных пунктах, контроля выхода абонента в радиоэфир и т.п. Технологически - это одни из относительно простых систем опознавания личности по речи.

Следующая задача, являющаяся основной в криминалистике, - это идентификация личности по речи, зафиксированной на спорной фонограмме. Близка к ней задача дифференциации реплик одной фонограммы по лицам без решения вопроса идентификации. Самая трудноразрешимая задача - это диагностика личности по речи.

При первом знакомстве с этими задачами может показаться, что они однотипны, т.е. если найти средство или способ решения одной из этих задач, то автоматически можно будет решать и любую другую. Однако при близком знакомстве и тем более при попытке создать обозначенные системы опознавания личности становится ясно, насколько сложен речевой механизм.

Криминалистическая фоноскопия не ограничивается только вопросами идентификации личности по речи. Это основной, решающий вопрос, но он практически всегда влечет за собой ряд других вопросов, среди которых чаще всего встречаются следующие:

- исследование признаков монтажа магнитной фонограммы;
- исследование признаков копирования фонограмм;
- идентификация магнитофона по магнитной фонограмме;
- диагностика и идентификация окружающей акустической среды звукозаписи;

- повышение разборчивости речевого сигнала, искаженного импульсными помехами или шумами.

Все эти задачи, за исключением последней, относятся к чисто физическим задачам, и их решение может быть полностью автоматизировано с помощью современных компьютерных средств.

Задача диагностики личности по речи в настоящее время корректно существует только в постановочной части. В своей же повседневной работе эксперты-фоноскописты, решая вопросы идентификации личности по речевым сигналам, постоянно вынуждены делать поправки на влияние “мешающих” факторов: алкогольное или наркотическое опьянение, функциональные расстройства речевого или голосового аппарата, изменчивость эмоционального состояния говорящего, различие речевых ситуаций и психологического настроя участников беседы, сознательное искажение речи или голоса говорящего с целью не быть правильно распознанным и др. Поэтому вольно или невольно, но любой эксперт так или иначе все же решает (хотя бы для себя, не отражая этого в заключении экспертизы) часть вопросов диагностики состояния говорящего. Это необходимо для того, чтобы в идентификационном исследовании правильно оценить степень значимости и достоверности индивидуальных признаков речевого сигнала. Накопление всего этого опыта дает надежду на возможное решение соответствующих диагностических задач. Однако если сравнивать возможности их решения с основной задачей идентификации личности по речи, то следует признать, что на настоящем этапе развития криминалистической фоноскопии надежды на эффективное решение диагностических задач почти призрачны.

Обычно, говоря о задаче криминалистической идентификации личности по речи, не заостряют внимания на содержательных или технологических нюансах ее постановки и решения. А в действительности они существуют и их так много, что можно говорить о разных технологиях идентификационных исследований. Выбор конкретной технологии диктуется следующими условиями:

- какие могут быть шумы и помехи, сопровождающие речевой сигнал, известны ли их характеристики или хотя бы класс;
- произносится ли речь отдельно от других высказываний или же на фоне других голосов;
- каков характер искажений, вносимых в речевой сигнал электроакустической аппаратурой;
- перемещается говорящий относительно микрофона или нет;
- говорит он непосредственно в микрофон (телефонную трубку) или находится на расстоянии, исключающем попадание воздушной струи на мембрану микрофона;
- произносятся слова слитно или раздельно;
- является ли речь логически связной или нет;

- произносится заранее заученный текст или же речь является спонтанной;
- совпадают или различаются тексты реплик опознаваемого лица в разных ситуациях;
- на одном и том же или на разных языках может говорить данное лицо в разных ситуациях;
- насколько различаются эмоциональные состояния опознаваемого лица в разных ситуациях;
- одинаково ли физиологическое состояние опознаваемого лица, нет ли существенных функциональных нарушений голоса и органов речеобразования;
- проявляется ли стремление опознаваемого лица сознательно исказить (замаскировать) свой голос, чтобы не быть правильно распознанным, или же, наоборот, говорящий готов “сотрудничать” с системой распознавания, чтобы быть правильно узнаваемым;
- каков психологический портрет опознаваемого лица, в какой мере проявляется в нем конформизм к голосу и речи собеседника;
- каковы требования к скорости реакции системы и надежности распознавания личности по речевому сигналу и др.

Данный перечень существенных условий функционирования системы опознавания личности по речевому сигналу, по-видимому, нельзя назвать полным. Наверняка существуют еще такие практические задачи, не попавшие еще в поле зрения проблемы опознавания личности по речевому сигналу, которые выдвинут еще ряд других требований, отличных от перечисленных.

Первоначально системы опознавания личности по речи разделялись на два класса: идентификации и верификации [70]. Их основной целью была организация “доступа по телефону к привилегированной информации” [4]. Б.Атал приводит более широкую классификацию систем опознавания говорящих: “Термин “опознавание дикторов” в статье относится к любому процессу принятия решения, который использует некоторые признаки речевого сигнала для определения, является ли именно данный человек диктором, произнесшим высказывание, в который включаются такие задачи, как идентификация, верификация, дискриминация и подтверждение дикторов” [4. С.49]. А.Розенберг [54] помимо перечисленных задач включает сюда и задачу классификации дикторов. Если в первых научных публикациях, посвященных проблемам опознавания дикторов, были, в основном, предприняты попытки дать классификацию систем и задач опознавания дикторов по методам их решения, то позже появились предложения классифицировать задачи по условиям их применения: на открытом или замкнутом множестве дикторов, по тексто-свободным или тексто-зависимым репликам [17].

В технической литературе задача идентификации личности по речи формулируется следующим образом: известно, что исследуемый речевой

сигнал принадлежит одному из N лиц, чьи образы речевых сигналов хранятся в памяти системы опознавания. Требуется определить, кому именно [47, 4, 49, 70]. Позже появляются модификации постановки этой задачи на открытом и закрытом множестве идентифицируемых лиц [17]. Задача верификации: “Определить, достаточно ли близок предъявляемый голос к его эталону и тем самым установить подлинность личности говорящего” [54].

Следует заметить, что иногда для обозначения одних и тех же задач вместо термина “верификация” используется “аутентификация” [86]. Это несколько запутывает задачи опознавания, поскольку, с одной стороны, оба эти термина обозначают проверку истинности, подлинности, установление достоверности, а с другой, - что, собственно, и вызывает путаницу среди задач опознавания личности по речи, актуальной является задача констатации преднамеренного изменения голоса [1]. В результате решения задачи требуется установить, является ли исследуемый речевой сигнал неискаженным (аутентичным) или же он преднамеренно искажен говорящим, чтобы не быть правильно опознанным. Таким образом, фактически имеются две прикладные задачи: определение искаженности голоса или речи говорящего (задача аутентификации) и отождествление личности по голосу или речи (задача верификации).

Задача дискриминации говорящих (иначе, дифференциации или различения говорящих) встречается в основном в криминалистике, когда необходимо по полилогу определить число лиц, участвующих в беседе, и какому лицу принадлежит каждая реплика. Фактически эта задача состоит из двух подзадач: определение числа лиц полилога и классификация реплик полилога по лицам.

Следует остановиться на различии задач идентификации говорящих, формулируемых в технической и криминалистической литературе. В криминалистической литературе под задачей идентификации личности по речи понимают решение вопроса: принадлежит ли речь, зафиксированная на двух разных фонограммах, представленных на исследование (экспертизу), одному или разным лицам. Такая постановка задачи не совпадает с задачей идентификации личности по речи, принятой в технической литературе. Поэтому и подходы к их решению, оценки точности и надежности (достоверности) выводов различаются. В криминалистике идентификация личности по речи осуществляется по максимально полному перечню идентификационных признаков и объединяет в себе исследование смыслового содержания речи, акустических свойств речевого сигнала и психологических аспектов речи. Настоящая работа имеет отношение только к изучению акустических свойств речевого сигнала, отражающих индивидуальные особенности работы органов речеобразования.

Если попытаться соотнести каждый конкретный тип научной задачи опознавания личности по речи к тем или иным нуждам практических служб правоохранительных органов, то в силу специфики деятельности последних

одни задачи оказываются более актуальными, другие - менее. Применительно к органам внутренних дел все задачи опознавания личности по речи оказываются примерно равноценными по актуальности:

а) задача криминалистической идентификации личности по речи наиболее актуальна на стадиях дознания и следствия, когда необходимо на основе имеющихся вещественных доказательств установить, кем конкретно были произнесены те или иные фразы. При этом в качестве вещественных доказательств (так называемых спорных фонограмм) используются фонограммы, представляемые, как правило, потерпевшими или свидетелями. Другую часть составляют официально изъятые копии фонограмм телефонных разговоров, зафиксированных в дежурных частях милиции и пожарной охраны, а также службе скорой помощи, справочных Аэрофлота. В качестве образцов обычно используются официально полученные образцы устной речи подозреваемых или процессуально оформленные фонограммы допросов, очных ставок с участием подозреваемых лиц;

б) задача установления числа лиц, участвовавших в полилоге, в явном виде обычно ставится после проведения некоторых оперативно-розыскных мероприятий. В этих случаях необходимо достоверно установить количество участников разговора и однозначно дифференцировать реплики по числу участников разговора. При этом никаких образцов речи у исследователей обычно не имеется. В неявном виде эта задача всегда присутствует при любой криминалистической идентификации личности по речи, когда, имея спорную фонограмму в форме полилога, необходимо сначала предварительно установить число участников полилога, закрепить каждую реплику за определенным участником, а затем уже, имея образцы речи подозреваемых, проводить идентификационное исследование;

в) задача аутентификации речи в явном виде ставится крайне редко. Однако любой эксперт-криминалист в каждом фоноскопическом исследовании так или иначе решает ее. Эта задача значительно переплетается с диагностикой алкогольного или наркотического опьянения, оценкой эмоционального состояния по речи и др. Состояние говорящих, чья речь зафиксирована на спорной фонограмме и на образцах устной речи, зачастую значительно различается. На спорной фонограмме можно обнаружить признаки эмоционального возбуждения, алкогольного или наркотического опьянения, а на образцах устной речи - далеко не всегда и не во всем названном комплексе. В этих случаях важны практический опыт и глубокие познания эксперта-криминалиста в области законов речеобразования.

Результаты решения перечисленных задач опознавания личности по речи используются в практической работе аппаратов дознания, уголовного розыска, БЭП, следствия и других смежных служб правоохранительных органов. Решение любой из задач в форме исследования или экспертизы проходит через экспертно-криминалистические подразделения, имеющие соот-

ветствующих специалистов и аппаратурную базу. Будем обобщенно называть их задачами криминалистической идентификации личности по речи.

§3. Информационные уровни фоноскопии

Вопросы, каковы, в какой форме и где фиксируются те самые материальные следы устной речи, по которым эксперт-криминалист производит идентификацию личности, не так просты, как это пытаются представить. Дело в том, что в отличие от почерка речевой сигнал в оригинале не виден и более того - весьма недолговечен. Изначально речевой сигнал - это всего лишь акустическая волна, излучаемая говорящим и несущая в себе фонетическую, смысловую, эмоциональную, индивидуальную и другую информацию. В информационном смысле - это богатейший источник.

Часто можно слышать утверждение о том, что речевой сигнал очень избыточен и его можно исказить как угодно до тех пор, пока не наступит, например неразборчивость. С этим можно согласиться, если исходить из отдельных целей и критериев, по которым производятся те или иные конкретные измерения информационной избыточности. Однако если речевой сигнал требуется исследовать всесторонне, например как в криминалистической экспертизе, то с тезисом об общей информационной избыточности речевого сигнала согласиться нельзя.

В этой связи очень интересны выводы, полученные Дж.Фланаганом [69] в результате анализа экспериментальных оценок скорости восприятия и передачи информации человеком. Он пишет: "Ни один эксперимент не показал, что человек способен обрабатывать информацию со скоростями, большими, чем примерно 50 дв.ед./с. Полагая, что эта цифра на самом деле представляет собой некоторую грубую оценку верхнего предела способности человека воспринимать информацию, человек может распорядиться этой своей способностью по-разному. Если, например, диктор достаточно быстро произносит случайный набор равновероятных фонем, слушающий может предъявить ко всему комплексу своих способностей требование правильно принимать письменный эквивалент различных звуков речи. При этом небольшую часть пропускной способности можно оставить для восприятия других характеристик речи, таких, как ударение, интонация, назализация звука, ритм и прочие свойства голоса. С другой стороны, если речь представляет собой праздную беседу с широкими статистическими связями и высокой избыточностью, слушающий может значительно большую часть своей пропускной способности выделить для анализа индивидуальных характеристик и особенностей артикуляции. В длительной беседе закономерности языка и достаточно эффективная память человека обычно дают возможность слушающему переключаться от расшифровки фонемного состава

к наблюдению за индивидуальными характеристиками и наоборот” [69. С.19].

Получается так, что если брать какой-то один критерий в отдельности, например скорость передачи текста высказывания, то при весьма широких возможностях передачи информации в акустическом диапазоне частот реально достигаемая скорость ее передачи в тексте составляет примерно 6 байт/с. Это очень низкий показатель. Однако нельзя забывать, что помимо фонетики в речевом сигнале отражаются и остальные аспекты речи: интонация, эмоции, функциональное состояние органов речеобразования, физиологическое, психологическое и интеллектуальное состояние говорящего. Поэтому если суммировать всю информацию, содержащуюся в речевом сигнале, то показатель будет весьма высоким. Каким - пока еще не известно. Возможно, что он окажется близким к верхней границе скорости передачи информации с помощью сигналов акустического диапазона частот.

Таким образом, было бы правильнее говорить не об информационной избыточности речевого сигнала вообще (как правило, в природе нет ничего избыточного), а о его высокой надежности, устойчивости и инвариантности к большому числу трансформаций.

Сказанное еще раз поясняет причину того, почему не удается создать универсальную систему опознавания говорящего, а появляются лишь некоторые специализированные системы, которые хорошо работают при условии, что изменяться и поэтому учитываться может только лишь какая-то часть факторов, определяющих речевой сигнал.

Наличие микрофона позволяет преобразовать (трансформировать) акустический сигнал в электрический и затем зафиксировать его на магнитном носителе в форме, весьма близко отражающей исходную форму акустической волны. Естественно, на магнитной ленте фиксируется не сам акустический сигнал, а его электромагнитный образ. Это и есть след речевого сигнала. Важно отметить, что магнитная фонограмма описывается точно теми же категориями следа, которые приняты в криминалистике.

Увидеть и исследовать речевой сигнал, зафиксированный на магнитной фонограмме, можно по-разному. Если есть необходимость явно исследовать непосредственно следы звуков, то нужно использовать специальные магнитооптические инструментальные средства, позволяющие полностью визуализировать дорожки звукозаписи на магнитной ленте и увидеть форму сигналов. Однако с помощью такого представления очень сложно охватить большие фрагменты сигнала и еще сложнее анализировать их содержание. Тем не менее такой способ визуализации следов иногда оказывается единственно приемлемым для решения некоторых вопросов криминалистической фоноскопии. Например, воспроизведение сигнала с ленты, подвергшейся механическим повреждениям, не позволяющим воспроизвести ее обычным образом на магнитофоне, а также идентификация магнитофона по оставляемым им трекам магнитных дорожек на ленте.

Обычный процесс воспроизведения следов звука с магнитной ленты по сути своей является обратным процессу его записи на ленту. Точно так же, как в процессе записи, лента протягивается мимо магнитных головок магнитофона и электромагнитная форма речевой волны трансформируется снова в акустическую, в результате чего громкоговорящие устройства излучают примерно такой же акустический сигнал, который записывался ранее на магнитную ленту. Этот наиболее доступный способ воспроизведения следа речевого сигнала является основным в экспертной практике. В такой форме обычно проводится аудитивный анализ содержания фонограмм. Чтобы произвести инструментальный анализ, электромагнитный сигнал с ленты поступает не только на громкоговорители, но и непосредственно на вход электроакустических измерительных приборов для измерения того или иного комплекса акустических параметров исследуемого речевого сигнала. Таким образом, неоднократно воспроизводя фонограмму, эксперт каждый раз может по-новому проанализировать один и тот же речевой сигнал: либо аудитивно (на слух), либо подавая его на специальные измерительные электроакустические приборы.

В зависимости от того, какая группа признаков сигнала исследуется, используются разные комплексы электроакустической аппаратуры. До внедрения цифровой технологии анализа сигналов приходилось для каждого отдельного вида криминалистической задачи применять разные типы регистрационно-измерительной аппаратуры. С переходом на компьютерную технологию анализа речевых сигналов аппаратный комплекс стал практически одинаковым для всех типов криминалистических задач: разные схемы регистрационных, измерительных и аналитических систем реализуются, образно говоря, в “мягкой” программной форме. То есть для проведения комплекса измерений акустических параметров исследуемого сигнала требуются не разные аппаратные системы, а разные системы программ, реализующие соответствующие измерения.

Проводить идентификационное исследование личности непосредственно по речевому сигналу практически невозможно. По этой причине некоторые юристы-процессуалисты, не вникая в природу речевых сигналов, могут сделать поспешный вывод о невозможности идентификации личности по речи, поскольку само базовое понятие криминалистики “след” оказывается весьма расплывчатым. Здесь нужно учитывать сложную природу речевого сигнала.

Действительно, звуковая волна в процессе своего распространения в воздухе, прохождения через системы преобразования в электромагнитную форму, линии передачи и звукозаписи могут так изменять свой вид, который внешне будет сильно отличаться от первоначального, имевшего место в момент произнесения непосредственно вблизи губ говорящего. Но это еще не означает, что искажается основная информация, которая первоначально присутствует в речевом сигнале. В большинстве случаев она остается неиз-

менной (исключение составляют случаи значимого искажения сигнала аппаратурой звукопередачи или звукозаписи либо дополнительными шумами и помехами). Органы слуха человека и предназначены для такой обработки звука, при которой все подобные виды изменения сигналов оказываются компенсированными. Иначе говоря, органы слуха извлекают из услышанного сигнала признаки, достаточно инвариантные ко всем реально встречающимся искажениям речевых сигналов.

На этом же основании строятся все системы идентификации личности по речевым сигналам: вместо анализа непосредственно акустической волны находятся признаки, инвариантные к тем преобразованиям и искажениям, которые всегда реально претерпевает речевой сигнал. До сих пор основными такими признаками, как правило, были всевозможные амплитудные спектры, которые инвариантны к сдвигу речевого сигнала во времени. Именно отражение динамики амплитудных спектров лежит в основе работы сонографа. На базе амплитудных спектров построено множество вторичных признаков, отражающих артикуляционные, голосовые и прочие характеристики работы речевого аппарата. Однако современные теоретические и прикладные исследования показывают, что это лишь одна часть признаков, а другая, не менее богатая, - скрывается в спектральных фазовых характеристиках речевого сигнала.

На анализе динамики амплитудных спектров речевых сигналов базируются практически все существующие криминалистические фоноскопические исследования. С их помощью определяются резонансные частоты речевого тракта при произнесении разных гласных и звонких согласных звуков речи. Эти резонансные частоты принято называть *формантами*.

Помимо статических значений формант при произнесении разных звуков большое значение в исследовании индивидуальных особенностей артикуляции имеет анализ динамики формант, которые впервые удалось увидеть только на широкополосных сонограммах. С помощью узкополосных сонограмм можно косвенно оценивать не только периоды работы голосовых связок, но и устойчивость периодичности их работы в разных ситуациях. С помощью совместного использования узкополосных и широкополосных сонограмм возможна оценка некоторых параметров формы голосового импульса. С помощью этих же типов сонограмм оцениваются и основные характеристики фрикативных и взрывных звуков речи. Таким образом, с помощью сонографа производятся разносторонние исследования всех звуков речи.

Основным предметом исследования в криминалистической фоноскопии является сам речевой сигнал. Поэтому для того, чтобы разработать эффективную компьютерную технологию идентификации личности по речевому сигналу, необходимо достаточно хорошо знать природу речевого сигнала. Однако на практике редко бывают случаи, когда перед экспертом ставится только один вопрос об идентификации личности по речи. Наиболее

распространенными являются сопутствующие вопросы по проверке достоверности фонограммы: выявить и описать признаки монтажа или копирования поступившей на исследование фонограммы, если таковые имеются. В этом случае эксперту требуются уже глубокие познания в области работы современных электроакустических приборов, используемых для улавливания, передачи и записи акустических сигналов на магнитную ленту. Для надежного решения задачи нахождения признаков монтажа фонограмм требуются предельно высокоточные методы оценки тех или иных электроакустических параметров, сигналов, фиксируемых на магнитной ленте в процессе записи или перезаписи речи на магнитофоне.

Не меньшие познания принципов и особенностей магнитной звукозаписи требуются при решении вопросов идентификации магнитофона, применявшегося для звукозаписи, диагностики типа использованного магнитофона, микрофона, магнитной ленты и т.п.

Дополнительные сведения требуются для решения вопросов диагностики и идентификации акустической среды, в которой производилась звукозапись исследуемой фонограммы. Здесь нужны знания из классической акустики для оценки реверберационных явлений в помещениях, диагностики и идентификации механизмов, приборов и прочих объектов и явлений живой и неживой природы, издающих определенные звуки.

Обобщая изложенное, можно еще раз подчеркнуть - проблема идентификации личности по речевому сигналу в общем виде чрезвычайно сложна и до сих пор она еще в принципе не решена. В настоящее время на практике с трудом решаются даже относительно “примитивные” задачи идентификации магнитофона по магнитной фонограмме, а идентификация личности во много раз сложнее, поскольку идентифицировать приходится по следам, отражающим не просто живой, но и интеллектуальный объект. Сложность заключается в том, что в речевом сигнале отражаются не только индивидуальные анатомические, физиологические и нейрофизиологические характеристики *индивидуума*, но и психологические, эмоциональные, интеллектуальные и иные особенности, определяющие конкретную *личность*. Кроме того, поскольку криминалисты обычно работают не с оригинальным речевым сигналом, а с его отражением на магнитной ленте (зачастую с копией этого отражения), то в исследуемом речевом сигнале одновременно отображается еще и множество электроакустических характеристик звукопередающей и записывающей аппаратуры.

Для построения компьютерной технологии фоновскопических исследований и экспертиз удобно выделить четыре информационно разных уровня анализа фонограмм речи:

физический (уровень акустического описания речевого сигнала). Этот уровень условно разбивается на три подуровня. На самом нижнем подуровне речевой сигнал описывается в терминах: амплитуда-время; на втором (подуровне параметрического описания) - в терминах: энергия в полосах

частот, высота и чистота голоса, форманты, спектральные переходы и т.п.; на третьем (подуровне модельного описания) - в терминах математического моделирования, вычисляются образы речевых сигналов. Последнее позволяет получить объективную картину артикуляционных и голосовых особенностей говорящего;

биологический (уровень анатомического и физиологического описания индивидуальных особенностей говорящего на базе объективных данных, полученных в результате физического анализа речевого сигнала);

психологический (уровень нейрофизиологического и психологического описания личности);

интеллектуальный (уровень лингвистического, интеллектуального, образовательного, профессионального и тому подобного описания индивидуальных особенностей говорящего).

Каждый из этих уровней расслаивается еще на ряд подуровней. Учитывая столь существенное информационное расслоение, во многих технических системах идентификации (верификации) по речи было бы правильнее говорить не об идентификации личности, а об идентификации индивидуума, поскольку личностные аспекты в них явно не присутствуют.

Как уже отмечалось, большое число криминалистических фоноскопических задач относится к уровню физического исследования акустических сигналов и самих магнитных носителей: исследования признаков монтажа, копирования, идентификации и диагностики магнитофонов, диагностики и идентификации акустической среды. Если же задача содержит вопросы исследования следов личности, то непременно криминалист сталкивается со всеми четырьмя названными выше уровнями информационного описания и анализа исследуемого объекта. Такой комплексный подход встречается лишь применительно к узкому кругу криминалистических исследований. Явно он возникает только в почерковедении. Из исследований, обычно к криминалистическим не относящихся, можно упомянуть идентификацию личности по способу совершения преступления. В большинстве же остальных традиционных видах криминалистических исследований выше второго информационного уровня (биологического) подниматься не приходится. Автороведческая экспертиза, наоборот, ниже третьего уровня (психологического) не опускается.

В идентификации личности по речевым сигналам высшие информационные уровни являются ведущими, так как описание речевых сигналов на нижних уровнях существенно зависит от сознательного управления речью и голосом. Иными словами, изменение речи на верхнем информационном уровне всегда ведет к изменению параметров, описывающих личность на более низком информационном уровне. Может возникнуть вопрос, как можно проводить идентификацию личности, если сознательное искажение голоса или речи может сместить (изменить) значения любого показателя более низкого уровня.

В какой-то мере его решению может помочь богатейший практический опыт проведения почерковедческих экспертиз, в которых найдены принципиальные решения, отражающие два подхода. *Первый* - это поиск аутентичных признаков, т.е. признаков, не искаженных в данный момент. В этом случае используется известный факт: чтобы человек ни предпринимал (пытался говорить другим голосом, деформировал свой речевой тракт, применял алкоголь, наркотики или иные медицинские препараты) - в каждом конкретном случае часть индивидуальных признаков остается неизменной. Всего речевых и голосовых признаков очень много [49, 52, 54], а сознательно можно контролировать лишь небольшое число одновременно. Таким образом, сила такого рода экспертиз - в поиске аутентичных признаков. Наибольший опыт в этом плане имеется у почерковедов [42].

Второй подход - ввести подозреваемого в состояние, аналогичное состоянию неизвестного лица на спорной фонограмме, добившись тем самым сопоставимости исследуемых речевых материалов. Для этого можно, например, попросить подозреваемого поговорить таким же голосом, как и на спорной фонограмме, или попытаться в разных вариантах исказить свою речь и т.п. Хотя этот путь практически реализуем с малой надежностью, о нем нельзя забывать. У этого подхода есть одна сторона, внешне кажущаяся слабой, дающая оппонентам пищу для отрицания правомерности такой процедуры, поскольку неспециалисту может показаться, что таким образом эксперт пытается подогнать образцы речи возможно невиновного подозреваемого лица под речевой сигнал неизвестного лица на спорной фонограмме. Однако глубокие научные исследования показали, что этого сделать невозможно. К такому выводу пришли многочисленные зарубежные и отечественные исследователи идентичности речевых сигналов имитаторов-профессионалов и речевых сигналов лиц, которых имитируют: всегда найдутся акустические параметры речевых сигналов, которые доказывают различия речи данного лица и его имитатора [47, 54].

Основным информационным уровнем криминалистической фоноскопии является физический. Именно он непосредственно связан с материальным носителем (вещественным доказательством) - магнитной лентой. На этом уровне не только воспроизводятся зафиксированные следы в виде речевых сигналов, но и измеряются любые его параметры, в том числе и те, которые говорящим не контролируются. Поскольку с помощью компьютерной технологии возможно измерение и даже физически неизмеримых параметров, то естественно возникает вопрос, возможна ли в будущем и каким образом может быть реализована полная автоматизация решения всех криминалистических фоноскопических задач.

Однозначно ответить на этот вопрос можно только в отношении полярных задач: "можно" - для задач идентификации магнитофона, исследования признаков монтажа или копирования фонограммы и т.п. (т.е. для

идентификации объектов неживой природы) и “нельзя” - для задач идентификации личности.

В отношении задач, лежащих между полярными информационными уровнями (физическим и интеллектуальным), однозначного ответа пока не предвидится. Может быть, с развитием принципиально новой идеологии построения компьютерной техники это станет возможным. Основные надежды в этой связи сейчас возлагаются на разрабатываемые нейрокомпьютеры, которые в определенной мере повторяют идею нейронного строения мозга. Проведенные испытания по распознаванию сложных объектов, в частности речи, с помощью нейрокомпьютеров дают хорошие прогнозы возможности автоматизированного распознавания личности по речи. В этом случае уже становятся сопоставимыми механизмы генерации речевых сигналов человеком и его анализа с помощью приборов, связанных с нейрокомпьютерами.

Невозможность полностью автоматического решения вопроса идентификации личности по речевому сигналу в настоящее время подтверждается еще и тем фактом, что любую существующую систему автоматического распознавания говорящего можно “обмануть” при желании быть неправильно распознанным. Особенно это может эффектно выглядеть, если известны те основные признаки речевых сигналов, на которых в системе распознавания строится решающее правило об идентичности говорящего. В этом случае достаточно лишь немного изменить свою речь, оставляя все остальные параметры, неконтролируемые системой распознавания, в их естественном диапазоне. Внешне это может показаться совсем незаметным. Этот фактор “сотрудничества” с системой распознавания является одним из решающих в системах автоматической верификации дикторов, которые эффективно применяются на всевозможных контрольно-пропускных пунктах, всевозможных системах ограниченного доступа к закрытой информации, финансовым операциям и т.п. В подобных системах весьма существенным является правило - хочешь быть правильно распознаваемым - принимай условие - говори в обычной своей манере примерно так, как делал это во время предоставления системе образцов речи.

Противоположной этой является ситуация распознавания произвольным образом искаженного речевого сигнала. Например, ускоренное (даже в два раза) воспроизведение записанного на магнитную ленту речевого сигнала без труда правильно распознается на слух, а опытным экспертом еще и правильно идентифицируется говоривший. Если провести здесь аналогию с письменной речью, то точно так же мы без труда читаем инструментально сжатый по любой из осей координат печатный или рукописный текст, а опытный эксперт еще и проводит по нему идентификационное исследование. Но приведенная аналогия не должна уравнивать достижения, возможности и особенности двух видов речевой технологии - устной и письменной. Технология устной речи еще весьма существенно отстает от письмен-

ной. Особенно ярко это можно проиллюстрировать путем сравнения двух одинаковых сторон той и другой речи. Например, автоматический синтез письменной речи, осуществляемый с помощью обычной пишущей машинки, прекрасно решен уже сотню лет назад (и сейчас эффективно реализован в виде компьютерного принтера), в то время как проблема удовлетворительного автоматического синтеза устной речи решается только сейчас, а еще несколько лет назад она была неразрешима.

§4. Технологии фоноскопических исследований

Все известные методики опознавания говорящего действуют по принципиально одинаковой схеме, описанной в той или иной форме во многих работах, посвященных проблемам опознавания личности по речи [4, 74, 49]. Общий вид этой схемы приведен на рис. 2.

Первым в работе системы опознавания выделяется этап получения образов речевых сигналов, инвариантных (или почти инвариантных) к изменениям, которые претерпевает речевой сигнал прежде, чем он поступит в систему опознавания. Если в системе опознавания имеется возможность свести к минимуму комплекс этих искажений (что, например, удастся сделать в системах верификации личности по речи на контрольно-пропускных пунктах), то построить систему достаточного описания образа речевого сигнала намного проще, чем сделать то же самое для случая большого числа искажающих факторов. В криминалистической практике классы искажающих факторов известны, но какого типа будут искажения в каждом конкретном случае, заранее предсказать нельзя. Уже одна только эта причина делает создание полностью автоматической системы опознавания личности, применимой в криминалистике, довольно проблематичным.

Результатом работы первого этапа являются два образа речевых сигналов: известной и неизвестной личностей. Заметим, что в некоторых системах по завершении первого этапа кроме образов речевых сигналов становятся известными и значения ряда факторов, искажающих исследуемый речевой сигнал (шумы, фон, реверберация, нелинейность искажений и т.п.). Знания о конкретных значениях или характере проявившихся искажающих факторов важны и используются на всех последующих этапах опознавания личности [74, 21].

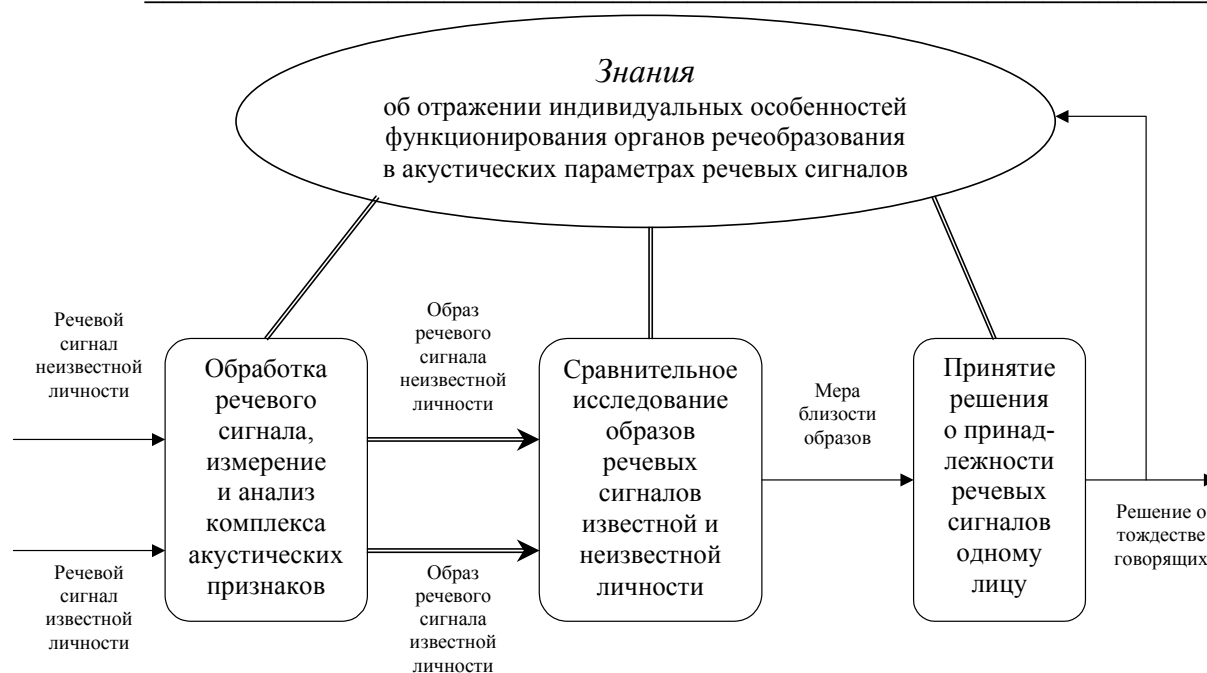


Рис. 2. Общая схема опознавания личности по речи.

На втором этапе производится сравнительное исследование всех построенных образов речевых сигналов известной и неизвестной личности. Если имеется адекватная модель описания поведения признаков, то могут производиться и оценки истинности тех или иных гипотез о принадлежности речевого сигнала одному и тому же лицу. В противном случае результатом сравнительного исследования являются оценки сопоставимости сравниваемых речевых образов с учетом искажающих факторов, установленных на предыдущем этапе исследования, и оценки меры близости полученных образов речевых сигналов неизвестной и известной личностей.

На третьем этапе формируется решение о принадлежности исследуемых речевых сигналов одному и тому же лицу. При этом используются все вычисленные меры близости исследуемых речевых образов с учетом знаний об их зависимости от искажающих факторов, установленных в конкретном идентификационном исследовании, и знаний, накопленных во всех подобных ранее проведенных исследованиях.

Первые системы опознавания по речи разрабатывались применительно к верификации личности на контрольно-пропускных пунктах к объекту ограниченного доступа. Дееспособность систем верификации основывается на гипотезе о том, что если человек сознательно не искажает свою речь, то фраза, произнесенная в привычной для него манере, имеет акустическое описание, отличное от акустического описания речевого сигнала другого человека. Эта гипотеза подтверждена многими исследованиями и в целом верна [47, 49, 4, 52]. Доказано также, что при включении в описание образа речевого сигнала достаточного набора информативных акустических при-

знаков система верификации без труда отличает речь одного человека от его имитатора [47, 54].

Сложность реализации системы верификации личности по речи связана с естественной вариативностью акустического описания речевого сигнала. Было найдено три способа явного или неявного учета этой вариативности.

Первый способ - это статистический анализ некоторого акустического параметра на всем протяжении речевого сигнала без различия характера сигнала на участках разных фонем. Так, были изучены и найдены некоторые признаки, отражающие индивидуальные особенности речи: средний спектр мощности сигнала, средняя частота основного тона и др. Анализ этих признаков [4, 76, 84] позволил создать не только экспериментальные установки на базе больших ЭВМ, но и начать серийный выпуск дешевых устройств верификации с использованием микропроцессорной техники [97]. Однако, несмотря на относительную простоту первого способа, он недостаточно эффективен при изменении эмоциональной насыщенности высказывания, форсировании или ослаблении фонации и т.п. [17]

Второй способ основан на построении параметрического описания речевого сигнала заданного текста и деформации получаемого описания так, чтобы нивелировать естественную вариативность длительности отдельных фонем и пауз. Для этого осуществляется незначительное нелинейное преобразование во времени параметрических описаний речевых сигналов верифицируемой личности, известной системе, таким образом, чтобы они были наиболее близки друг к другу. Характер и степень допустимой деформации параметрического описания определяются той естественной вариативностью, которая имеет место в речи личности, известной системе. В итоге решение о тождестве говорящих принимается не на основе абсолютного совпадения исходного параметрического описания, а на основе почти совпадающей динамики параметров речевого сигнала. Основой такого сопоставления является метод динамического программирования [47]. Подобный метод был использован в системе автоматической верификации личности на фирме Bell Laboratories [47, 54].

Несмотря на то, что во втором способе верификации происходит явный учет естественной вариативности параметрического описания речевого сигнала, он не нашел широкого применения и дальнейшего развития из-за отсутствия гибкости системы к различию произносимых текстов и довольно сложной (относительно других систем верификации) системы обработки и анализа речевых образов.

Третий способ, который условно можно назвать методом опорных точек, нашел широкое применение и продолжает развиваться во многих системах опознавания (не только верификации) личности. В контрольно-пропускной системе, разработанной фирмой Texas Instruments, автоматическая верификация осуществляется по отдельным опорным точкам парамет-

рического описания речевого сигнала. Эти точки выбираются в областях максимальной энергии речевого сигнала на участках гласных звуков [54]. При этом в памяти системы хранятся эталоны параметрического описания отдельных гласных звуков слов в парольной четырехсловной фразе.

Аналитическая часть процесса верификации начинается с того, что каждый эталон отдельного звука размещается на вновь поступившем в систему параметрическом описании речевого сигнала верифицируемой личности. Размещение эталонного описания гласных звуков осуществляется по критерию минимизации функции ошибок. Точка положения эталона, доставляющая минимум функции ошибок, указывает место опорной точки в поступившем речевом сигнале. Дальнейшее сравнение речевых образов производится по значениям параметрического описания речевого сигнала в опорных точках анализируемой фразы и соответствующего параметрического описания эталонов. Такой подход позволяет избавиться от вариативности параметрического описания речевого сигнала во времени, не решая все вопросы учета естественной вариативности речевого сигнала. Однако сам принцип позволил развить подход опознавания говорящих по опорным точкам и для текстуально-различающихся высказываний.

Следует заметить, что при кажущейся простоте и завершенности системы верификации личности эта задача в широком смысле все еще остается нерешенной. Сложность ее заключается в том, что удостовериться в тождестве удастся лишь при условии, что говорящий “сотрудничает” с системой верификации и желает быть отождествленным с лицом, чей речевой эталон хранится в системе. Однако не всегда одного желания быть правильно признанным оказывается достаточно. Иногда (применительно к нетипичному для данного лица эмоциональному состоянию) параметрическое описание речевого сигнала изменяется настолько сильно, что система верификации принимает решение об отсутствии тождества верифицируемого лицу, речевой образ которого в системе имеется.

Из-за возможности сотрудничества говорящего с системой верификации ее широкое внедрение затруднено. В криминалистической практике помимо всех перечисленных причин подобные системы верификации не находят применения еще и потому, что они требуют совпадения аппаратурных и акустических искажений речевого сигнала на этапе построения эталонного образа и на этапе построения образа верифицируемой личности. На практике это осуществить крайне трудно. Однако гибкость третьего способа верификации по опорным точкам и новые подходы в выделении ключевых признаков позволили разработать ряд систем опознавания, применимых и в криминалистике. Среди них выделяются две системы: SASIS (Semi-Automatic Speaker Identification System) и СКИФ (Система Криминалистического Исследования Фонограмм).

В основу полуавтоматической системы SASIS, разработанной еще в 1975 г., был заложен принцип, аналогичный принципу опорных точек. При-

влекательность системы SASIS заключается в том, что она максимально соответствует принятой в криминалистике технологии идентификационных исследований личности, поэтому на ее описании следует остановиться несколько подробнее.

Порядок работы системы SASIS в общих чертах следующий [78, 100]. Оператор прослушивает и записывает слова, содержащиеся в речи неизвестного на спорной фонограмме и в образцах речи подозреваемого, и готовит их фонетическую транскрипцию. Сравнивая фонетические транскрипции, оператор отбирает одинаковые трезвучия таким образом, чтобы необходимый для сравнения звук находился в середине. В перечень звуков речи, используемых для сравнительного исследования, входят носовые и гласные звуки, всего 13 фонетических элементов [100]. Определив перечень анализируемых трезвучий, оператор вводит фрагменты речи, содержащие эти фонемы, в память ЭВМ.

После ввода речевых сигналов оператор вызывает программу обработки сигналов и построения сонограмм на графическом терминале. Длительность фрагментов, отображаемых на сонограмме, составляет 1,1 с. Анализируя сонограммы, оператор выбирает с помощью визирной линии на экране графического дисплея участки речевого сигнала с теми фонетическими элементами, которые предстоит включить в дальнейшее сравнительное исследование. При этом правильность выбора фонемы по сонограмме контролируется аудитивно путем воспроизведения коротких (длительностью от 100 до 300 мс) фрагментов речевых сигналов в окрестности визирной линии. Во время прослушивания микроучастков на экране дисплея отображается форма воспроизводимого сигнала. Удостоверившись в правильности позиционирования микроучастка, оператор выбирает временное окно таким образом, чтобы в него входило ровно три периода сигнала. Это позволяет в дальнейшем получать более точное спектральное представление сигнала, используя преобразование Фурье, синхронно периоду основного тона.

После вычисления 30 акустических параметров всех выбранных фонем выполняется программа сравнения групп выбранных фонем каждого из 13 классов в речи неизвестного на спорной фонограмме и известного лица на образцах его речи. В результате работы этой программы получают оценки идентичности акустических признаков речевых сигналов идентифицируемых лиц и оценки вероятностей ошибок идентификации.

Сильной стороной системы SASIS является ее технологическая близость к принятой в криминалистике общей схеме идентификационных исследований по этапам:

- предварительного исследования с оценкой пригодности и сопоставимости сравниваемых объектов;
- отдельного исследования объектов с выявлением групповых и индивидуальных признаков;
- сравнительного исследования индивидуальных признаков;

- подготовки решения с указанием его надежности.

Слабой стороной системы SASIS, не позволившей ей получить должного развития и внедрения, является значительный уклон в сторону автоматизации последних этапов идентификационного исследования и принятия решения о тождестве лиц, в то же время наиболее трудоемкие операции первой части исследования полностью отданы оператору. Среди них - операции, связанные с аудитивно-визуальной сегментацией речевого сигнала по фонемам и с выбором временного окна для повышения точности спектрального оценивания. Такой перекося, когда рутинные трудоемкие операции по повышению точности вычисления акустических параметров переносятся не на вычислительную технику, а отдаются оператору, в то время как аналитическая деятельность по сравнению результатов измерений и подготовке решения с учетом степени сопоставимости образцов речи и речи на спорной фонограмме полностью переданы вычислительной системе, в целом неприемлем.

В этом плане противоположностью системе SASIS является система СКИФ, разработка которой была начата Институтом систем управления Республики Грузия [3], а завершена ее автором - В.Д.Сердюковым. В системе СКИФ все трудоемкие операции по сегментации и классификации фонем отданы измерительно-вычислительной технике, с помощью которой производится не только автоматическое измерение параметров сигнала, но и по заданным правилам строится образ речевого сигнала.

В образ речевого сигнала включается последовательность “опорных ключевых сегментов ударных гласных звуков” [3], которые выделяются из речевого сигнала по критерию максимума общей энергии сигнала. Каждый сегмент описывается фиксированным набором параметров: спектр мощности, общая мощность, частота основного тона и др. Параметрическое описание речевого сигнала формируется на выходе специального анализатора сигналов.

Формируемый в системе образ речевого сигнала, основанный на оригинальном методе нормализации речевых сигналов [58], в итоге оказывается более устойчивым к большому классу шумов, не зависящим от текста, темпа и типа высказывания. Это достигается благодаря тому, что в исследуемом речевом сигнале основными объектами анализа выступают самые мощные и менее всего подверженные коартикуляционным искажениям ударные гласные звуки. Система СКИФ ориентирована на работу с речевыми сигналами в диапазоне частот до 4 кГц. В настоящее описание образа речевого сигнала одного лица в системе занимает около 2 Кбайт памяти, что позволяет на базе типового персонального компьютера вести и хранить фонокартотеку нескольких десятков тысяч лиц.

Система идентификации личности по речи, развиваемая в Украинской академии МВД, строится в основном на исследовании спектральных переходов речевого сигнала [18]. Важным преимуществом спектральных пере-

ходов является их независимость от линейных амплитудно-частотных искажений, что достигается благодаря вычитанию логарифмических амплитудных спектров смежных участков речевого сигнала. Недостатком спектральных переходов является их чувствительность к шумам. Кроме того, в спектральных переходах отражается довольно ограниченный набор индивидуальных особенностей речеобразования. Технологические недостатки системы [18] в целом примерно те же, что и у системы SASIS - это большой объем работы по аудитивной сегментации речевого сигнала на фонетически значимые элементы.

В ЭКЦ МВД России методика идентификации личности по речи строится на параллельном акустическом и лингвистическом анализе речи. Методика опознавания личности по лингвистическим признакам речи разработана во ВНИИ МВД СССР в 1985 г. [74] и используется на практике. Методика исследования акустических, в частности фониатрических, признаков еще не разработана в достаточной мере из-за отсутствия приемлемого по точности и гибкости применения инструментария акустического анализа речевых сигналов в криминалистических целях. Финансируемая ЭКЦ МВД России разработка компьютерной системы малого предприятия "Центр речевых технологий" под названием SIS (Sound Identification System) по своим функциональным возможностям повторяет некоторые известные системы подобного класса. Однако она уступает им по таким важнейшим для криминалистики показателям, как разрешающая способность измерения и анализ разного рода акустических признаков. В этом плане изделие ПАКОРС (Программно-аппаратный комплекс очистки речевых сигналов), промышленно выпускаемое АО "Дальняя связь", значительно превосходит как по техническим характеристикам, так и по функциональным возможностям, в первую очередь необходимым именно для криминалистической фоноскопии.

Разработка элементов компьютерной технологии фоноскопических исследований ведется и в Академии МВД России. На основе проведенных теоретических исследований, учета экспертного опыта фоноскопических исследований и видения принципиальных путей решения вопросов фоноскопии, изложенных в настоящей работе, в разработке компьютерной системе SV (Signal Viewer) основное внимание уделяется высокой точности, достоверности, разносторонности и функциональной гибкости исследования фонообъектов. Ряд основных разработанных элементов этой компьютерной фоноскопии будет изложен ниже.

§5. Компьютерная технология фоноскопических исследований

Говоря о технологии той или иной деятельности (не обязательно производственной), будем придерживаться следующего определения этого понятия. Технология - это комплекс инструментальных средств и система методов их применения.

Компьютерную технологию фоноскопических исследований и экспертиз можно определить как комплекс аппаратных и программных средств обработки и анализа звуковых сигналов для решения частных экспертно-криминалистических вопросов и задач.

Средства фоноскопии разделяются на аналоговые и цифровые (компьютерные). Аналоговые средства - это микрофоны, усилители, магнитофоны, фильтры, эквалайзеры, акустические системы, наушники, аналоговые измерительно-регистрирующие приборы (самописцы уровня сигнала, сонографы, генераторы, источники питания и т.п.) и многое тому подобное. В данной работе проблемы использования или разработки этих средств рассматриваться не будут, поскольку это выходит за рамки исследования. Но изредка, естественно, будет констатироваться потребность их применения без рассмотрения вопросов модификации, разработки или совершенствования. То же самое касается и второй стороны технологии - методов их применения.

С одной стороны, данная работа ограничивается только компьютерными аспектами фоноскопических исследований и экспертиз. Но охватить их все также весьма проблематично из-за ограничений по объему изложения. Поэтому из всего множества проблем компьютерной технологии большее внимание будет уделено лишь одной сфере - математическому и программному обеспечению, принимая вторую сторону компьютеров - современное аппаратное обеспечение - как данное и оставляя ее вне рамок исследования.

С другой стороны, программно-математическая сторона компьютерной технологии в чистом виде также рассматривается не как самостоятельный объект или предмет исследования, а как средство реализации предлагаемых новых методов автоматизированного цифрового исследования сигналов.

Поэтому в настоящей работе основное внимание уделяется теоретическим исследованиям и разработкам новых методов цифровой обработки и анализа звуковых сигналов, необходимых для решения задач фоноскопических исследований и экспертиз. На их разработку, естественно, существенное влияние оказывают перечисленные элементы современной компьютерной технологии: аппаратная часть компьютеров, их программное обеспечение, новые языки, технологии программирования и многое другое.

Реализация предлагаемых разработок видится в двух формах: программной и аппаратной. Но поскольку последняя требует очень тщательной и долговременной проверки на практике, чтобы не упустить важные нюансы, которые после реализации их в “металле” корректировать и исправлять практически невозможно, то, как и все современные компьютерные технологии, они сначала будут ориентироваться в основном только на программную реализацию.

Подытоживая изложенное, отметим, что как заданное в настоящем исследовании будут считаться:

- множество современных аппаратных (аналоговых и компьютерных) средств обработки и анализа сигналов;
- способы перехода от континуального представления звуковых сигналов к дискретному;
- методы и алгоритмы быстрого преобразования Фурье;
- методы автоматического управления и многое другое.

Опираясь на эту базу и другие современные достижения науки и техники, будет предпринята попытка создания системы новых специфичных средств и методов, составляющих технологию фоноскопических исследований и экспертиз.

Рассматривая понятие технологии, нельзя обойти вопрос определения понятия объекта фоноскопических исследований. Ниже этот вопрос будет рассмотрен более подробно. Здесь лишь отметим, что под *фонообъектом* будет пониматься реальный объект, генерирующий и излучающий в звуковом диапазоне частот сигналы, которые преобразованы в цифровую форму и хранятся в памяти компьютера в виде отдельных файлов. Под категорию фонообъекта попадают не только реальные объекты, излучающие звук (человек или механизм) или видоизменяющие его (резонирующая, реверберирующая среда), но и множество электроакустических приспособлений (магнитофоны, микрофоны, генераторы, выпрямители источников питания) и просто сигналов. Поэтому при исследовании объекта, описываемого некоторыми сигналами звукового диапазона частот, не будут устанавливаться жесткие границы. И чтобы в исследования входили объекты или явления, оставляющие о себе следы в виде сигналов широкого диапазона частот, будет полагаться, что они могут иметь частотные компоненты и ниже 20 Гц и выше 20 кГц, но не выше 200-300 кГц (чтобы оставить в поле зрения анализ, например, частоты подмагничивания в магнитофоне 70-90 кГц).