

Глава 5. КОМПЬЮТЕРНАЯ ТЕХНОЛОГИЯ РЕШЕНИЯ ЧАСТНЫХ ВОПРОСОВ ФОНОСКОПИИ

§1. Исследование артикуляции

На основе исследования артикуляционных признаков речи построены практически все современные системы автоматической верификации диктора. Основная причина тому - простота получения сглаженных во времени широкополосных амплитудных спектров речевых сигналов, получаемых чаще всего не цифровыми методами, а с помощью аналоговых приборов (электроакустических фильтров). При этом используются значения уровней мощности речевого сигнала в небольшом числе частотных полос, обычно не превышающем 30-ти. В этом случае получается несколько огрубленная оценка, в которой могут отсутствовать реально имеющиеся тонкие детали треков резонансных частот исследуемого диктора. Для криминалистики желательно иметь лучшую по разрешающей способности картину исследуемого речевого материала. С этой целью предлагается использовать амплитудные сонограммы с предельно высоким разрешением по частоте и времени. Таковые получаются с помощью описанных выше методов расчета спектров и отображения сонограмм.

Имея сонограмму высокого качества, эксперт в принципе применяет те же самые методы, которые реализуются в автоматизированных системах верификации дикторов. Например, автоматическая сегментация речевого сигнала с помощью фонетической функции Пирогова с наименьшей точностью и надежностью выполняется экспертом визуально по сонограмме. Но помимо этого опытный глаз эксперта всегда улавливает массу индивидуальных особенностей, которые чаще всего не отражаются в слишком уж широкополосных, да еще к тому же и сглаженных амплитудных спектрах или в параметрах, построенных на основе производных этих спектров. Поэтому в экспертной практике пока что предпочтительнее более осторожное обращение с речевым сигналом, близкое, например, к подходу, реализованному в системе SASIS.

Большая часть идентификационных признаков строится на базе сопоставления нормативного и исследуемого реально имеющегося произношения тех или иных звукосочетаний. К сожалению, общепризнанного атласа таких нормативов сейчас пока еще нет. И поэтому эксперт в подобных случаях вынужден более опираться на результаты фундаментальных и при-

кладных научных исследований [4, 8, 10, 13, 16, 17, 23, 30, 33, 34, 37, 38, 50, 52, 54, 58, 61, 62, 67, 69], а также свой практический опыт и опыт коллег.

Приводимые ниже примеры графических образов тех или иных фоно-объектов были получены с помощью компьютерной системы “Signal Viewer”, в которой реализованы вышеизложенные методы компьютерной фоноскопии [21].

Для иллюстрации на рис. 27 приведены примеры одних и тех же трезвучий типа гласный-согласный-гласный (слово “алло”), произнесенных разными дикторами, из которых у первых двух артикуляционные признаки можно считать близкими к нормативным. Особенно это касается динамической структуры амплитудного спектра на всем участке смычно-проходного звука “л”. Внимательно изучая сонограммы, опытный эксперт всегда найдет в них те или иные особенности, которые после оценки их устойчивости на разных участках речевого сигнала могут перейти из индивидуальных особенностей в разряд идентификационных артикуляционных признаков речи (см. рис. 28).

Из приведенных примеров видно, что бывают случаи резкого отличия артикуляционных признаков у разных дикторов. В этих случаях наиболее сложны ситуации сегментной неразличимости отдельных участков фонем. Так, на рис. 29 приведены примеры сонограмм двух дикторов, у которых очень трудно установить границы фонемы “л”. Это происходит, в частности, и потому, что используемая амплитудная сонограмма отражает лишь половину исходной информации, а вторая половина (фазовая) остается за кадром. Установить границы этой фонемы может помочь посегментное прослушивание исследуемой сонограммы. Для этой цели эксперт имеет возможность выделить произвольный участок сонограммы и прослушать его содержание.

Прослушивание отдельных выделенных сегментов особенно ценно при анализе сигнала в шумах и помехах, когда воспроизводят смежные или отдельные фонемы и может быть даже не всю фонему, а только какую-то ее хорошо сохранившуюся часть, не искаженную, например, мощной импульсной помехой. Последняя возможность анализа весьма важна, поскольку отдельные виды помех, располагаясь в том или ином месте фонемы или рядом с ней, могут в корне изменить реальное восприятие речевого сигнала, полностью изменив его смысл или сделав его полностью неузнаваемым (или непонимаемым).

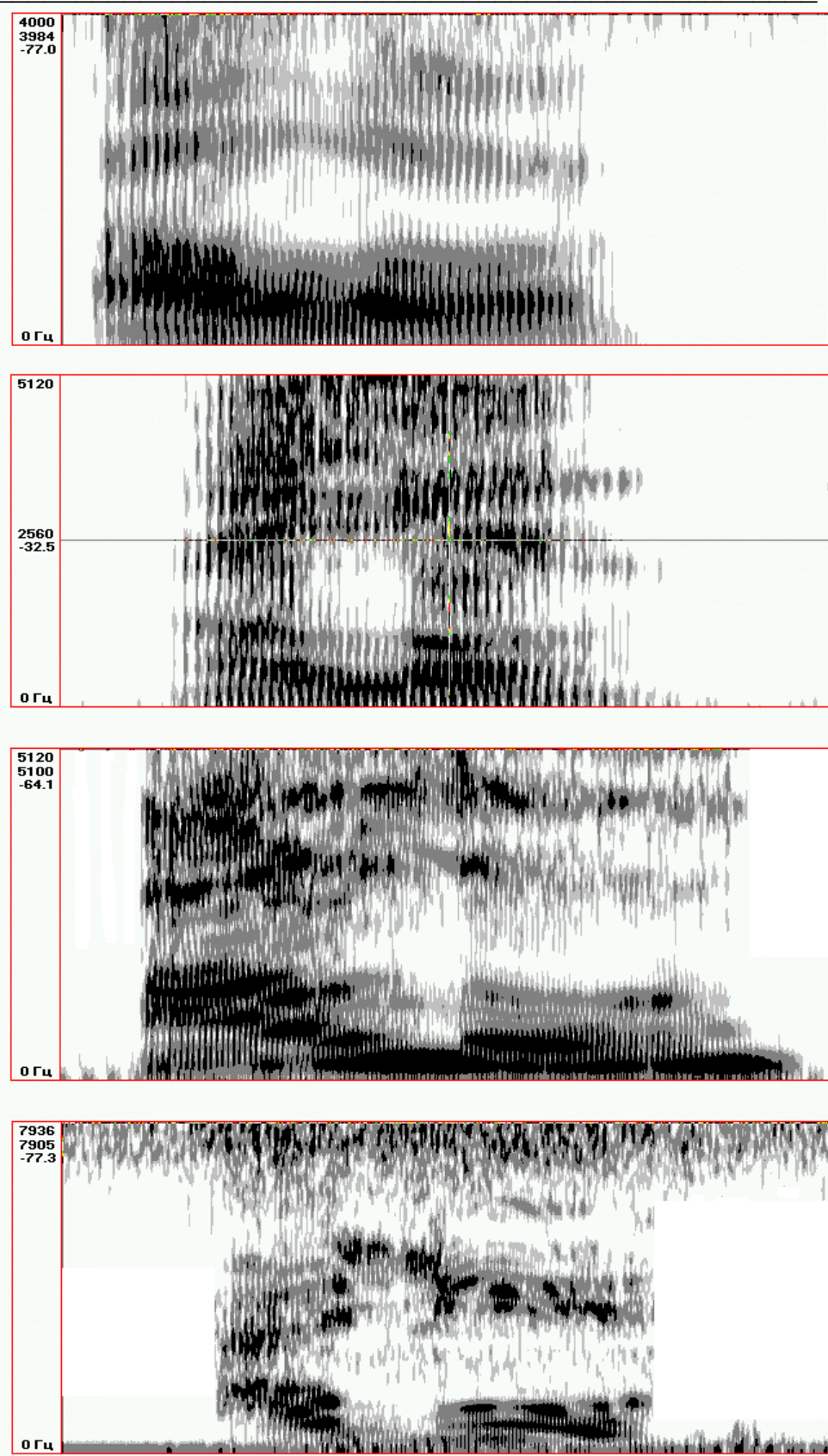


Рис. 27. Широкополосные сонограммы слова “алло” разных лиц с произношением, близким к нормативному

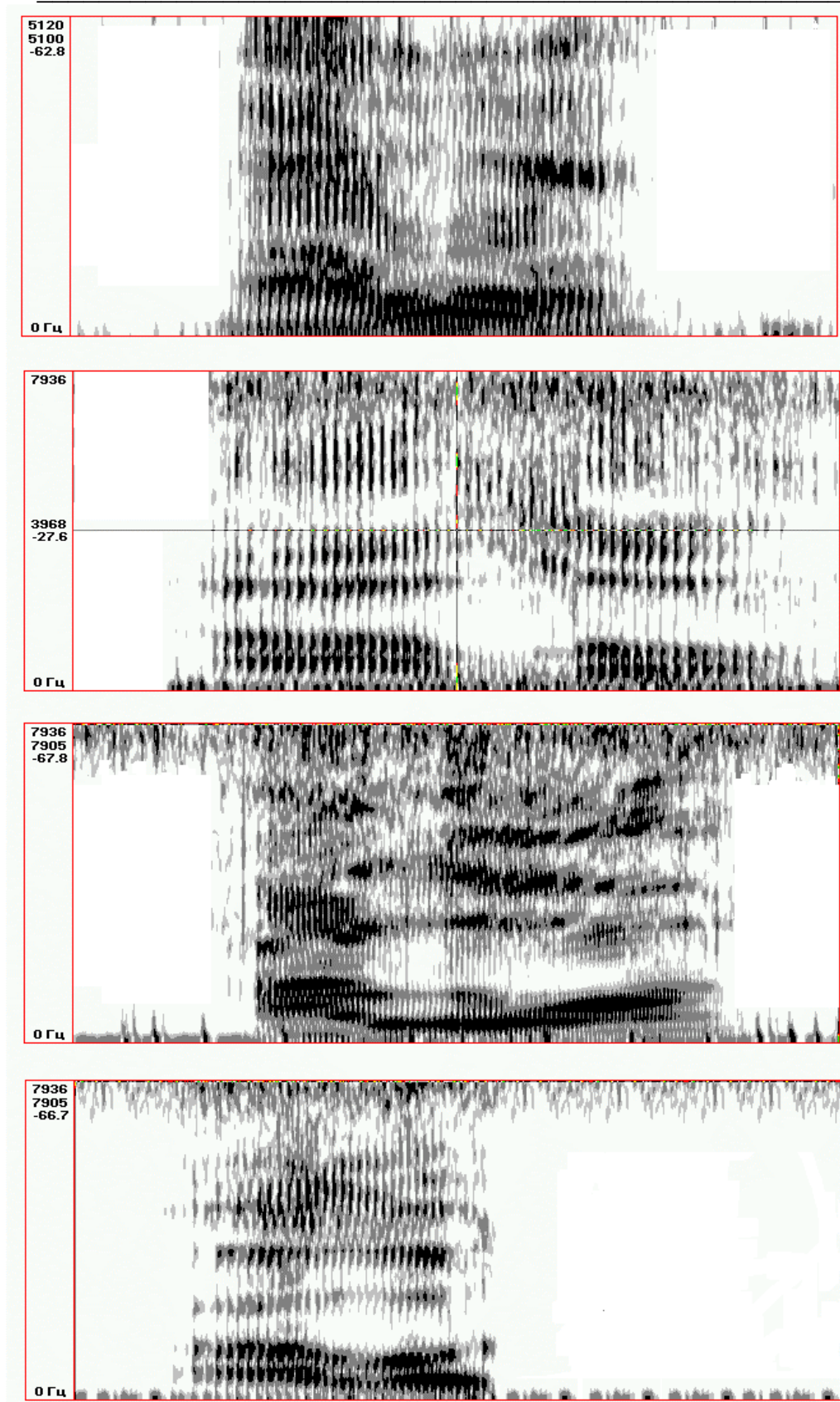


Рис. 28. Широкополосные сонограммы слова “алло” разных лиц с хорошим произношением

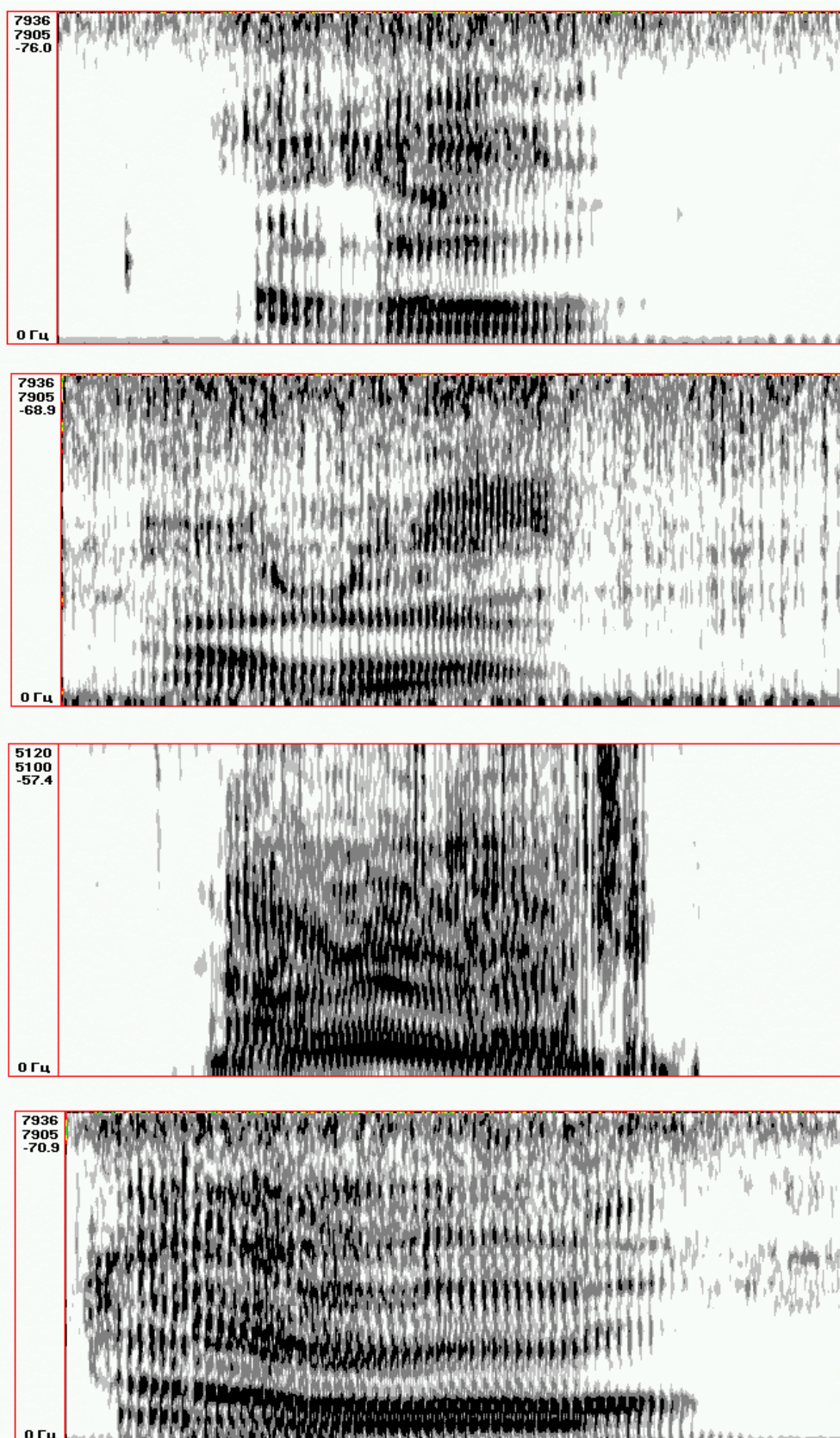


Рис. 29. Широкополосные сонограммы слова “алло” разных лиц с произношением, далеким от нормативного

В рассмотренном ранее случае (см. рис. 19) резонансные частоты речевого тракта расположены настолько плотно друг к другу и имеют близкие

значения уровней мощности, что на сонограмме они практически неразличимы. Увидеть их более четко помогают гармонограммы. На рис. 19 приведены сонограмма и гармонограмма такого же слова “алло”, что и на рис. 27-29. С их помощью удастся различить большое число резонансных частот речевого тракта (заведомо больше двух) в зоне, где обычно предполагается обнаруживать первые две форманты.

§2. Исследование голоса

Ниже приводятся некоторые предварительные результаты исследований, полученные в процессе испытания дее-способности разработанных методов исследования голоса. Объем этих исследований еще слишком мал, чтобы делать окончательные выводы об универсальности обнаруженных особенностей тех или иных элементов образов речевых сигналов. Но эта задача и не ставится перед настоящим исследованием. Представляемая разработка дает лишь средство обнаружения и анализа индивидуальных особенностей голоса в речи. Исследование устойчивости во времени, частоты встречаемости и информативности тех или иных акустических признаков речевых сигналов, применяемых в решении криминалистических задач опознавания личности, должно проводиться отдельно учеными-криминалистами.

Элементарные статистики

Исследование любых признаков речевых сигналов существенно усложняется тем, что сам речевой сигнал или его параметрическое описание нельзя однозначно отнести ни к классу детерминированных, ни случайных процессов. В действительности проявляются оба эти полярные качества. Поэтому естественно применение статистических методов анализа, которые составили основу всех представляемых ниже результатов.

Для решения задач опознавания личности по речи актуальной является проблема определения исчерпывающей системы статистик (по А.Н.Колмогорову [14]) описания законов изменения голоса в речи. В некоторых работах приводятся попытки описать статистическую устойчивость частоты основного тона с помощью минимального числа статистик. Например, в [44] предлагается описывать закономерность поведения частоты основного тона с помощью всего лишь двух статистик: среднего и дисперсии логарифма частоты основного тона. Однако, как показывают результаты многих исследователей [6, 66, 51, 98], описать поведение частоты основного тона голоса в речи только этими статистиками невозможно, поскольку хорошо известны случаи бимодальности плотности распределения частоты

основного тона, что уже требует, как минимум, системы из четырех статистик.

Вопрос “От системы скольких параметров зависит функция распределения частоты основного тона голоса в речи данного лица в данном конкретном случае” в настоящее время не решен. И следует признать, что он не решен не только практически, но и теоретически. Интуитивно ясно, что разные эмоциональные и физиологические состояния говорящего, разные типы речевого поведения и т.п. по-своему влияют на функцию распределения частоты основного тона голоса в речи. Однако каково множество параметров, описывающих функцию распределения (не говоря уже об их содержательной стороне), а следовательно, каково минимальное множество статистик, необходимое для полного описания закономерностей поведения частоты основного тона в данном речевом сигнале, пока не известно. Поэтому на первых этапах исследования индивидуальных особенностей поведения частоты основного тона голоса в речи можно считать оправданным включение в исследование максимального числа всевозможных статистик, отсеивая в процессе дальнейших исследований те из них, которые оказываются сильно зависимыми от факторов, оценить влияние и воздействие которых практически невозможно.

В настоящем исследовании в список анализируемых были включены следующие элементарные статистики [27] частоты основного тона голоса:

- минимальное и максимальное значение;
- среднее значение;
- среднеквадратичное отклонение;
- третий и четвертый центральные моменты;
- процентиля от 1 до 99%;
- коэффициент вариации;
- моды функции плотности распределения.

В качестве экспериментальных использовались магнитные фонограммы рассказов автобиографий разных лиц. Речевые материалы частично содержали общие слова (“родился”, “году”, “я” и т.п.), но в целом были текстуально различными. Рассказы осуществлялись без предварительной подготовки. Эмоциональная окраска каждого рассказа была нейтральной (типичной) для данного лица.

С помощью этих фонограмм определялось, какой объем выборки частоты основного тона необходим для того, чтобы та или иная выборочная статистика была устойчива к изменениям текста высказывания. При этом минимальный объем выборки определялся величиной, превышение которой приводило к тому, что значения выборочной статистики начинали принадлежать одному и тому же интервалу, полуширина которого равна точности измерения отдельного значения частоты основного тона. Оценка доверительных интервалов истинных значений статистических параметров не про-

изводилась, поскольку характер распределения параметров частоты основного тона в разных конкретных случаях может быть разным, и каким именно - неизвестно.

У разных лиц объемы выборок для получения устойчивых значений выборочных статистик оказались разными. И, кроме того, выборочные значения разных статистик проявляли разную устойчивость. Лучшую устойчивость проявили: среднее значение, среднеквадратичное отклонение и средние процентиля (40-60%). В среднем по исследованным лицам эти выборочные статистики оказываются устойчиво постоянными, если объем выборки превышает 8 с чистого времени звучания голоса в речи (определение объема выборки частоты основного тона удобнее указывать не в количестве отсчетов дискретного по времени параметрического описания речевого сигнала, а более универсально - в единицах времени, поскольку в первом случае приходилось бы помимо объема выборки в единицах отсчета постоянно напоминать и о длине шага временного дискрета параметрического описания). Устойчивые значения статистик: коэффициент вариации, третий и четвертый центральные моменты, а также крайние значения процентилей (около 10 и 90%) - получались в том случае, когда продолжительность времени звучания голоса в анализируемом речевом сигнале превышала 12 с.

После превышения объема выборки 12-15 с устойчиво проявляются моды плотности распределения частоты основного тона. Число и расположение этих мод у разных лиц проявляется по-разному. Для примера на рис. 30 показаны выборочные плотности распределения частоты основного тона голоса двух мужчин Ш. и Л. с очень близкими средними значениями частоты основного тона - 100,1 Гц и 102,5 Гц. Подобная ярко выраженная полимодальность плотности распределения частоты основного тона наблюдалась не у всех обследованных лиц. У отдельных лиц устойчиво проявлялись лишь одна-две моды. В плотности распределения производной контура частоты основного тона устойчиво проявлялась только одна мода (см. рис. 31). Подробнее об анализе полимодальности будет сказано ниже, а здесь лишь приведем описание замеченных закономерностей поведения некоторых простейших статистик.

Сопоставление устойчивых выборочных статистик современных высказываний (произнесенных с интервалом в несколько минут) показало полное совпадение всех статистик за исключением: минимального и максимального значения частоты основного тона и процентилей, близких к 100%.

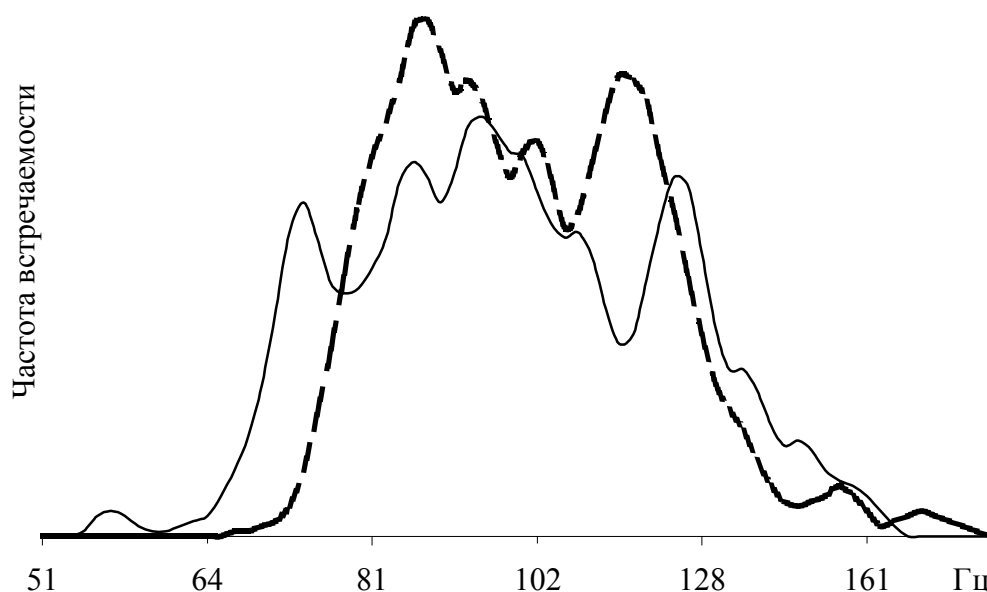


Рис. 30. Распределения частот основного тона голоса лиц “Ш” и “Л”

Для исследования статистик частоты основного тона несовременных высказываний использовались записи живых бесед нескольких лиц друг с другом, выполненных с интервалом во времени не менее 10 дней (но до одного года). В результате выяснилось, что более половины из всех исследованных статистик не сохраняли свои значения.

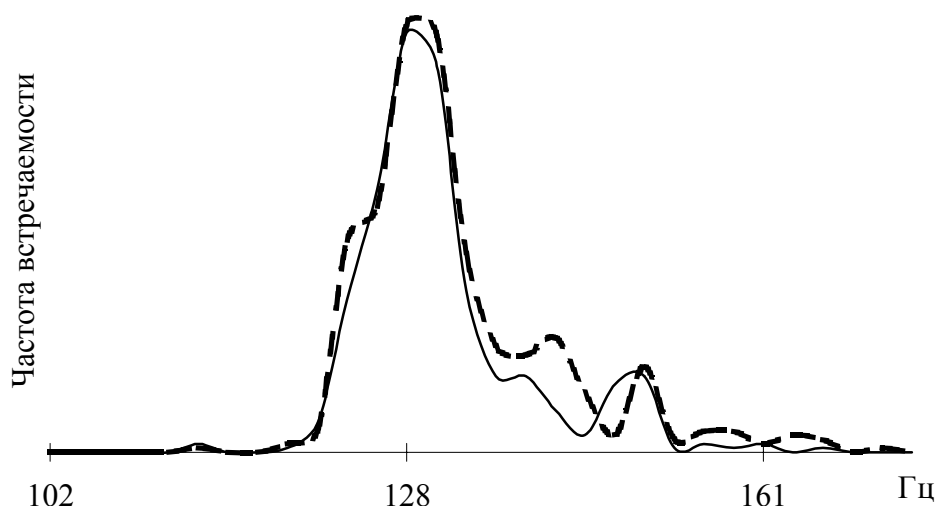


Рис. 31. Квазиунимодальность распределения частоты основного тона одного и того же лица, но в разное время

Так, у одного из обследуемых (в этом исследовании принимало участие восемь человек) в двух различных диалогах, в которых явно проявилось различие эмоциональных состояний (в одном случае - легкое возбуждение, а в другом - некоторая успокоенность), средняя частота основного тона составила соответственно 193 Гц и 173 Гц. Причем в обоих случаях речевое поведение было примерно одинаковым по таким показателям, как

громкость, темп речи и типы предложений. С ростом средней частоты основного тона в относительно более возбужденном состоянии возросло и среднеквадратичное отклонение. Из всех статистических параметров наиболее смещенными оказались процентиля - выше 50%. Эти результаты хорошо согласуются с данными, опубликованными в [11].

У всех обследованных чаще всего неизменными оставались лишь самые нижние процентиля. Обычно в речи, близкой к эмоционально нейтральной, неизменными оставались процентиля около 10% и ниже. А при возрастании эмоциональной насыщенности речи свои значения сохраняли только самые нижние процентиля - около 2-5%. Но и они иногда смещались вверх. С одной стороны, это может быть объяснено относительно большой неустойчивостью процентилей, близких к нулю, а с другой - иногда обнаруживавшимся явным смещением нижних процентилей вверх, если на протяжении более 10-12 с общего времени звучания голоса в речи основной тон находился все время на верхних или средних частотах и почти не опускался на нижний уровень ненапряженной спокойной фонации. Такое явление имело место в эмоционально сильно возбужденной речи.

В качестве иллюстрации приведем результаты измерения выборочных функций распределения частоты основного тона голоса двух мужчин "К" и "Б" в разных речевых ситуациях. На рис. 32 показаны две выборочные функции распределения частоты основного тона в речи лица "Б" в диалогах с другими лицами, разделенных во времени двумя неделями. Объемы выборки частоты основного тона в демонстрируемых распределениях составили 14 и 16 с. Характерными субъективными особенностями поведения лица "Б" во время обоих экспериментов являются ощутимая скованность и эмоциональная подавленность. Функции распределения частоты основного тона в обоих случаях очень близки друг к другу в области процентилей ниже 50% и особенно близки в области процентилей 2-10%.

На рис. 32 показаны примеры выборочных функций распределения частоты основного тона в речи другого лица "К". Разница по времени между исследуемыми речевыми материалами лица "К" также составляет две недели. Объем выборки частоты основного тона составил 15 и 21 с. Субъективно эмоциональные состояния лица "К" в двух случаях полярно противоположны: возбужденность (агрессивность) в одном случае и угнетенность (подавленность, пассивность) - в другом. Функции распределения частоты основного тона в данных ситуациях существенно различны. Однако хорошо видно (см. рис. 32), что и в этом случае значения самых нижних процентилей (2-4%) совпадают.

Неизменность самых нижних процентилей распределения частоты основного тона (2-10%) наблюдалась у всех восьми обследованных в случае эмоционально спокойной речи. Получить устойчивые и неизменные в течение длительного времени оценки иных элементарных статистик, указанных в начале данного раздела, не удалось.

Приведенные результаты экспериментов основаны на малой совокупности испытуемых. Однако одну качественную особенность голоса в речи, обнаруженную у всех обследованных, следует подчеркнуть - наиболее устойчивыми к изменению речевой ситуации, эмоционального состояния во времени являются низшие процентиля в окрестности 2-10%. Или, иначе говоря, ширина диапазона голоса в речи данного лица может сильно варьироваться от ситуации к ситуации, но нижняя граница этого диапазона (нижние процентиля) относительно устойчива, а расширение диапазона голоса происходит за счет поднятия верхней границы (верхние процентиля) диапазона

голоса. Поэтому в криминалистической практике применение такой статистики, как средняя частота основного тона, довольно проблематично. Но и исключать эту статистику из криминалистических исследований нельзя, поскольку, например, еще не установлена возможная взаимосвязь между изменением среднеквадратичного отклонения и среднего значения частоты основного тона. И, в частности, еще не проверена гипотеза о совпадении средних значений частоты основного тона данного лица в разных речевых ситуациях при условии совпадения среднеквадратичных отклонений частоты основного тона.

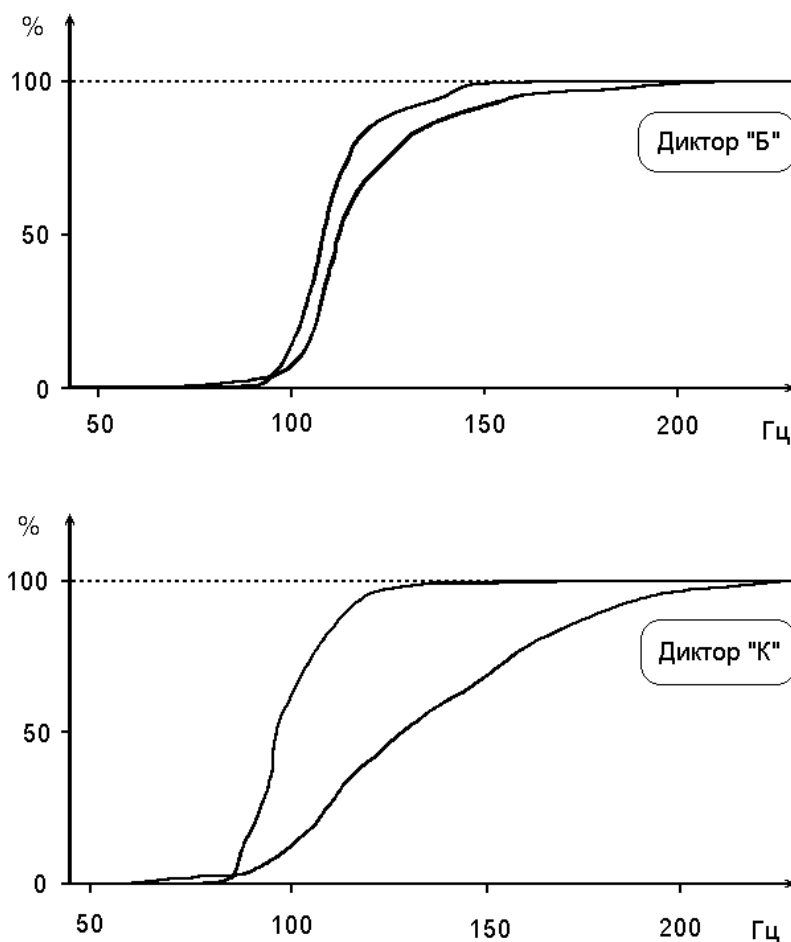


Рис. 32. Функции распределения частоты основного тона дикторов в разное время

Голосовые доминанты и методы их измерения

В научных публикациях [50, 51, 44, 6, 98], содержащих анализ плотности распределения частоты основного тона голоса в речи, приводятся три разных типа плотности распределения, которые наблюдались в тех или иных исследованиях: нормальное, логарифмически-нормальное и бимодальное. Причин такого расхождения может быть много: это и способ отбора участников эксперимента, и принятое в эксперименте речевое поведение, и виды речевых сообщений, и технология получения образцов речи, и метод измерения частоты основного тона, и многое другое. Но основной причиной, по-видимому, является сама испытуемая личность, точнее, свойственный данному лицу индивидуальный характер распределения частоты основного тона в той или иной речевой ситуации.

Простой визуальный анализ контуров частоты основного тона в координатах (частота)*(время) показывает, что у одних лиц контур сильно изрезан, в нем преобладают участки резкого подъема, а участки звучания голоса с постоянной частотой и участки плавного уменьшения с асимптотическим приближением к нижней частоте практически отсутствуют. У другого лица, наоборот, контуры очень плавно меняют вид. В них могут непрерывно чередоваться подряд несколько подъемов, задержек на некоторой частоте и плавных спадов. У третьих лиц обнаруживается поразительное постоянство высоты голоса на всем протяжении длительного высказывания (очень монотонная речь). Поэтому естественно ожидать, что у таких трех разных типов поведения контуров частоты основного тона плотности распределения частоты основного тона тоже будут отличаться по виду.

К сожалению, сопоставить результаты исследований плотности распределений частоты основного тона, приведенные разными авторами, очень сложно, а иногда и просто невозможно. Невозможно, потому, что если авторы и указывают в общих чертах способ измерения частоты основного тона, то не указывают, каким образом в их исследовании производилась дифференциация речевого сигнала на участки, где отражается работа голосовых связок (тональные участки речевого сигнала), и участки, на которых голосовые связки не работают. Решение этого, казалось бы, простого вопроса в действительности так же сложно, как и измерение частоты основного тона. А может быть, и сложнее, поскольку если с помощью того или иного метода формально можно производить измерение частоты основного тона, в результате чего получать некоторое допустимое значение, то кроме этого должна еще даваться и оценка степени надежности того, что измерена частота (период) именно интересующего явления - периодичности сигнала в целом, а не какого-либо иного явления, например, периодичности следования экстремумов речевого сигнала.

Произвести аудио-визуальную сегментацию речевого сигнала на тональные - нетональные участки, а затем измерить частоту основного тона довольно просто и по сигналограмме речевого сигнала (достаточно чистого и частотно неограниченного), и по сонограмме его узкополосного спектра. Сделать то же самое автоматически принципиально сложно.

В настоящем исследовании применена следующая система выделения участков речевого сигнала, на которых обнаруживается явление периодичности сигнала.

Как уже отмечалось выше, максимум амплитуды функции индикатора основного тона отражает степень гармоничности структуры узкополосного спектра, и поэтому может служить мерой степени периодичности речевого сигнала в данном временном окне расчета узкополосного спектра. Для удобства дальнейшего изложения будем условно называть амплитуду функции индикатора основного тона показателем периодичности или звонкости звука. Последнее название оправдывается тем, что на участках гласных и звонких согласных звуков амплитуда функции индикатора основного тона значительно выше, чем на участках глухих согласных звуков.

Выбрав некоторое пороговое значение показателя звонкости, можно произвести автоматическую сегментацию данного речевого сигнала на тональные и нетональные участки.

В настоящем исследовании порог выбирался оператором, которому предоставлялась гистограмма распределения показателя периодичности сигнала, с помощью которой предлагалось указать то значение показателя, которое должно считаться гранью в классификации периодичности-непериодичности сигнала. После этого строилась условная (определяемая указанным порогом) выборочная функция плотности распределения частоты основного тона данного речевого сигнала.

Прежде чем перейти к описанию результатов анализа выборочных распределений частоты основного тона, отметим одну принципиальную особенность предложенной классификации точек параметрического описания речевого сигнала по критерию периодичности. На рис. 33 представлены фазовые портреты частоты основного тона одного и того же слова “мыло”. При непрерывном произнесении этого слова контур частоты основного тона также непрерывен. Но при выборе разных порогов показателя периодичности отдельные участки траектории контура становятся отнесенными к классу непериодических. В частности, отметим, что при достаточно высоком значении порога показателя периодичности в фазовом портрете остаются лишь те участки, где производная частоты основного тона близка к нулю. И это вполне закономерно, поскольку на участках резкого изменения частоты основного тона, т.е. на участках с большой производной контура частоты основного тона, речевой сигнал явно аperiodичен (в математическом понимании периодичности функции).

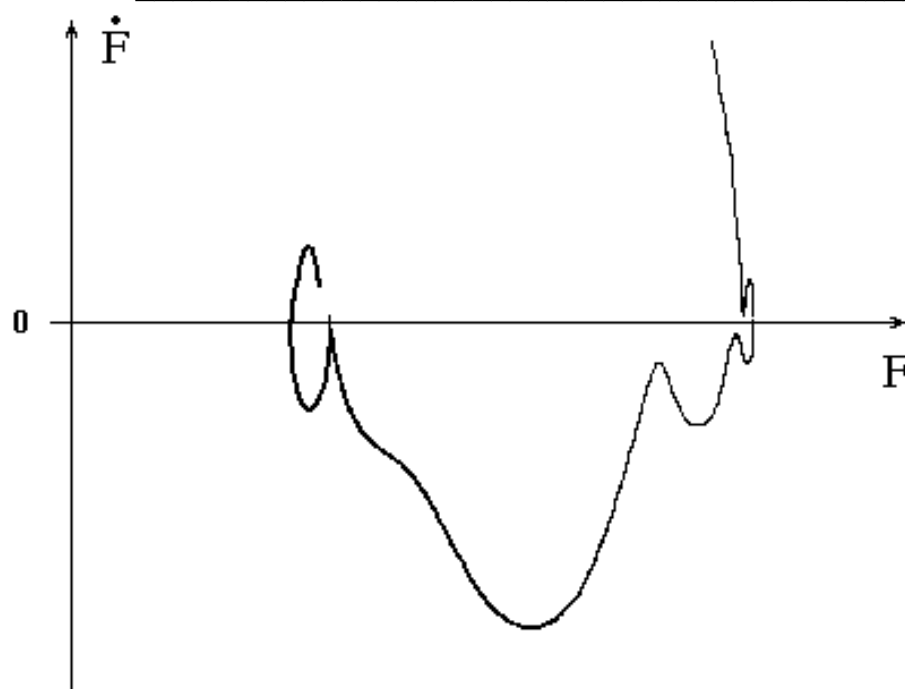


Рис. 33. Фазовый портрет частоты основного тона голоса при произнесении слова “мыло”

Построение выборочных функций плотности распределения частоты основного тона с разными пороговыми значениями критерия периодичности восьми обследованных лиц показало, что моды функции плотности распределения у разных лиц в разных ситуациях проявляются

по-разному. Однако замечена одна общая для всех лиц закономерность: глобальная мода распределения частоты основного тона в эмоционально спокойной речи проявлялась и в относительно возбужденной речи. Но при этом она обычно переставала быть глобальной. Некоторые моды, близкие по высоте к глобальной в эмоционально спокойной речи, находились в выборочных распределениях основного тона этого же лица во время других испытаний. При этом высоты мод менялись непредсказуемо. Некоторые моды иногда исчезали, а некоторые, ранее не проявлявшиеся, возникали с относительно большой высотой. При всей подвижности формы распределения большое число мод, проявившихся в отдельных испытаниях, совпадало с модами распределения частоты основного тона голоса, построенного на большой обучающей выборке. Анализ этого явления показал, что решающий вклад в образование мод вносят квазипостоянные, относительно длительные участки траектории частоты основного тона. Такие участки в обычной речи встречаются довольно часто, и их длительность может варьироваться от 0,04 до 1 с и более (при заполнении речевой паузы звуками типа “Э-э”).

Была предпринята попытка построить выборочную условную плотность распределения частоты основного тона - плотность распределения значений частоты основного тона в тех точках речевого сигнала, где огибающая контура основного тона квазипостоянна. С этой целью для автоматического выделения квазипостоянных участков контура использовались два критерия: первый - это описанный выше способ выделения тональных участков речевого сигнала, а второй - близость к нулю производной оги-

бающей контура частоты основного тона. При этом производная огибающей контура вычислялась путем линейной аппроксимации параметров частоты основного тона во временной окрестности 50 мс методом наименьших взвешенных квадратов.

На рис. 34 приведены две выборочные условные плотности распределения частоты основного тона лица М. в разных речевых ситуациях, разделенных во времени приблизительно одним месяцем. На верхнем рисунке показана плотность распределения, полученная по большой выборке из речевого сигнала, в котором общая длительность тональных участков превышала 19 с, а на нижнем - составила около 5 с. Из семи мод (совпадение числа мод в приведенных примерах случайно) в этих плотностях распределений неизменными остаются четыре моды: А, В, Е и Ж. Причем глобальная мода В на верхнем рисунке проявилась и на нижнем, но перестала быть глобальной. То же самое относится и к моде А на нижнем рисунке.

Подобное явление имело место у всех восьми испытуемых. Возвращаясь к проблеме изменчивости средней частоты основного тона голоса у одного лица в несовременных высказываниях в разных речевых ситуациях, можно утверждать, что в голосе человека более устойчиво проявляются узкие области доминантных частот, в которых траектория частоты основного тона находится чаще всего. И в зависимости от того, что, как и в каком эмоциональном состоянии говорит данная личность, в плотности распределения лучше проявляются те или иные доминантные частоты основного тона голоса.

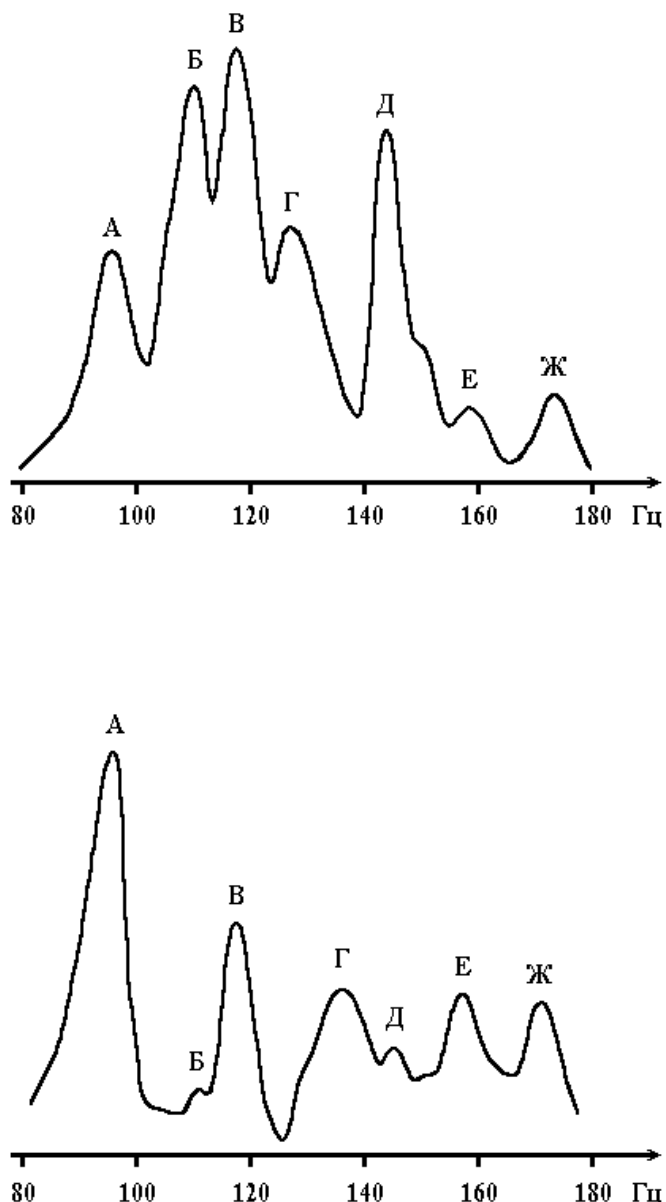


Рис. 34. Выборочные функции распределения частоты основного тона одного и того же лица в разных эмоциональных состояниях

Далее будем называть проявление нескольких мод в плотности распределения частоты основного тона голоса в речи явлением полимодальности распределения частоты основного тона [19].

Причины проявления полимодальности распределения частоты основного тона голоса пока еще не ясны, но некоторые гипотезы (не более того) могут быть высказаны. Так, одной из причин бимодальности может быть хорошо известное в вокальном искусстве наличие двух основных голосовых регистров - грудного и головного [56], из которых нижний - грудной - может быть разбит еще, по крайней мере, на два регистра [29]. Другой причиной полимодальности может быть наличие у каждого человека нижней (базовой) доминантной частоты основного тона голоса, относительно которой более высокие частоты формируются под контролем системы слухового восприятия говорящего, производящей свою индивидуальную оценку благозвучности голоса на разных частотах и оценку характера голосовых переходов с одной частоты на другую. Существование нижней (базовой) доминантной частоты косвенно подтверждают эмпирические вышеприведенные данные об устойчивости к изменениям эмоционального состояния говорящего 2-10% процентилей функции распределения частоты основного тона его голоса.

Следует отметить, что явление полимодальности распределения голоса в речи, обнаруженное разработанными и описанными выше инструментальными методами, ранее было известно. Так, в работе [29] мы находим: “Слушая начинающего певца, педагог находит у него обычно, на ряду с неудовлетворительными, тускло и резко звучащими звуками, один или несколько уже более или менее хорошо звучащих. В то время как неблагополучные звуки производятся с резко выраженным усилием, обрываются, сопровождаясь гримасами, напряжением лица, шеи, а иногда и всего тела, тона хорошо звучащие издаются певцом относительно легко, без особого напряжения ... Эти, как я их называл бы, основные звуки наименее поддаются дурным привычкам и навыкам, приобретаемым певцом-самоучкой. Если задать себе вопрос, на каких именно звуках диапазона можно у начинающего певца легче всего выявить наиболее приемлемые модификации звучания его голоса, то окажется, что таковыми в большинстве случаев являются два или три звука его диапазона, лежащие преимущественно в середине первой октавы его голоса - чаще всего те самые звуки, которые являются для него наиболее привычными в повседневной разговорной речи” [29. С.133].

Приведенное качественное описание особенностей голоса, известное более полувека назад, не объясняет причины полимодальности распределения частоты основного тона, но подсказывает, на какие качественные особенности голоса следует обратить внимание, чтобы получить новые количе-

ственные акустические характеристики, отражающие индивидуальные особенности голоса в речи. Остановимся на утверждении о том, что тонкий музыкальный слух различает в голосе “неудовлетворительно и хорошо звучащие тона”. Из приведенной выдержки [29] видно, что в понятие хорошего звучания включаются такие характеристики, как степень легкости длительной фонации звука определенной высоты и относительная чистота звучания голоса (“тускло или хорошо звучащие звуки”).

Легкость длительной фонации отдельных звуков, по-видимому, и приводит к тому, что в речи на этих частотах траектория основного тона задерживается чаще и дольше всего. А поскольку именно такие участки огибающей траектории частоты основного тона включались в выборочные плотности условного распределения, то можно предположить, что описанные выше методы: а) качественного определения на слух “основных звуков” и б) инструментального определения “голосовых доминант” - по сути своей близки друг к другу.

Но на слух оценивается не только способность длительной фонации звука на неизменной высоте, но и “тусклость” его звучания. Если поставить задачу: оценить акустическими методами степень чистоты звучания голоса, то первая трудность, с которой сразу приходится сталкиваться, - это отсутствие общепринятого определения понятия “чистота звучания голоса”. Выше было предложено оценивать меру чистоты звука по высоте максимума функции индикатора основного тона в окрестности значения периода основного тона.

Усредняя функции индикатора основного тона по всем участкам речевого сигнала, на которых обнаруживается полезная спектральная составляющая в области частот до 640 Гц, удастся получить функцию, подобную выборочной функции плотности распределения частоты основного тона. Эти две функции имеют следующие различия:

а) средняя функция индикатора основного тона представляет распределение периодов основного тона во временной области, а плотность распределения частоты основного тона - в частотной;

б) в выборочной плотности распределения частоты основного тона каждое отдельное значение основного тона включается в распределение с единичным весом, независимо от степени периодичности сигнала, а в средней функции индикатора основного тона каждое отдельное значение оценки периода основного тона входит с весом, пропорциональным ширине диапазона частот, в котором проявилась гармоническая структура узкополосного спектра, что, как уже отмечалось, является одной из мер оценки степени периодичности сигнала.

Таким образом, средняя функция индикатора основного тона позволяет выделять в голосе те периоды (или высоты звучания голоса), которые чаще встречаются в работе голосовых связок и речевой сигнал на которых максимально периодичен. Но, как уже отмечалось ранее, чистота звучания

голоса зависит от степени постоянства периодов следования голосовых импульсов и от степени постоянства их формы. Нарушение этого постоянства приводит или к дрожанию голоса и неустойчивости его звучания, или к проявлению сипоты. Поэтому в средней функции индикатора основного тона отражаются обе характерные индивидуальные особенности голоса, описанные в [29]: и более частое нахождение высоты голоса, и более чистое звучание голоса на отдельных частотах.

В заключение прокомментируем различия между формами плотностей распределения частоты основного тона голоса в речи, приводимых в указанных выше публикациях и продемонстрированных в данной главе. Из множества причин этих различий выделим три, на наш взгляд, основные:

- 1) различие метода измерения частоты основного тона голоса в речи;
- 2) различие в оценке меры периодичности сигнала в точке измерения частоты основного тона;
- 3) различие в учете веса отдельных выборочных значений в плотности распределения частоты основного тона.

Точность измерения частоты основного тона по первому обертону (в спектральной или временной области) в речи, когда речевой сигнал можно лишь условно считать стационарным (что предполагается в подобных подходах), снижается из-за того, что формантная структура звука может существенно сместить обертоны от истинного значения гармоник. Поэтому при измерении частоты основного тона голоса на коротких по длительности звучания участках разных гласных звуков, произносимых с практически постоянными периодами следования голосовых импульсов, может возникнуть разброс измеренных значений частоты основного тона, характер которого зависит от произносимых звуков и конкретно применяемого метода измерения частоты основного тона.

Вторая и третья причины обусловлены тем, что в разных системах измерения и анализа частоты основного тона на переходных участках разных звуков по-разному определяется критерий достоверности (надежности) измеряемого значения частоты основного тона. Например, на участках переходов звонких смычных, смычно-проходных и гласных звуков происходят значительные вариации расстояний между экстремумами речевого сигнала. И эти вариации, связанные с резкими амплитудно-фазовыми изменениями тональных участков речевых сигналов, в некоторых системах, основанных на методе измерения периода основного тона во временной области [87], автоматически считаются вариациями периодов и включаются в распределение частоты (или периода) основного тона с одним и тем же весом. А в быстрой речи общая длительность подобных переходных участков сравнима с общей длительностью квазистационарных участков речевого сигнала. Поэтому большая доля измеренных значений частоты основного тона оказывается недостоверной, многие оценки частоты основного тона оказываются смещенными в ту или иную сторону в зависимости от кон-

кретного перехода между отдельными звуками. При этом все эти оценки включаются в выборочные распределения с постоянным весом, что приводит к «размыванию» выборочной плотности распределения частоты основного тона голоса, измеренного по речевому сигналу, по сравнению с истинной плотностью распределения частоты основного тона.

Рассмотрим некоторые качественные характеристики голоса, которые без особого труда обнаруживаются аудитивно и которые можно количественно анализировать.

Одной из сложнейших проблем криминалистической идентификации личности по речи является диагностика аутентичности звучания голоса [2]. Очень часто при отборе образцов речи, не желая быть распознанным, подозреваемый начинает изменять свой голос: или выходя за нижнюю либо верхнюю границы своего диапазона высоты звучания голоса, или начиная добавлять хрипоту в звучание голоса. Аудитивно это иногда удается диагностировать. Имеется ли возможность проводить такую диагностику инструментально? Пока что такие инструментальные методы не существуют [2]. Однако из личной экспертной практики автора известно, что любой выход частоты основного тона за верхнюю границу речевого диапазона данного лица приводит к тому, что, пытаясь постоянно держать голос выше, например, на целую октаву, в начале и (или) в конце законченного высказывания контуры частоты основного тона имеют характерный изгиб (с большим абсолютным значением производной) в сторону нижних частот. Особенно хорошо это проявляется в конце повествовательного высказывания, когда в нормальной речи при окончании речевого сигнала относительно долгим гласным звуком наблюдается плавное снижение частоты основного тона с асимптотическим приближением к некоторой постоянной частоте. В сознательно искаженной высоте голоса путем перехода в верхний частотный диапазон наблюдается не плавное асимптотическое снижение к некоторой постоянной частоте, а резкое (с большой отрицательной производной) снижение с внезапным обрывом траектории частоты основного тона без выхода ее на квазистационарный уровень.

Имея техническое средство точного измерения и отображения огибающей контура частоты основного тона, по указанным признакам можно производить диагностическое исследование вероятности нахождения частоты основного тона не в своем обычном частотном диапазоне, а за его верхней границей.

Сложнее обстоит дело с выходом голоса за нижнюю границу своего обычного диапазона. В этом случае с помощью узкополосных сонограмм до сих пор не удавалось обнаружить признаков этого явления. Здесь может оказаться полезным свойство функции индикатора основного тона отражать меру стабильности работы голосовых связок.

Исследуя поведение показателя периодичности P_{136} вокальных голосов у лиц, не являющихся певцами, было обнаружено, что на нотах,

сознательно взятых ниже нижней границы речевого диапазона данного лица, голос звучит очень неровно. Наблюдаются значительные резкие скачки в контуре частоты основного тона. И, что особенно следует подчеркнуть, средний уровень показателя периодичности сигнала ниже на нотах, выходящих за нижний речевой диапазон, и выше на нотах, лежащих в речевом диапазоне высоты звучания голоса испытуемого лица.

Другая проблема, которая до сих пор еще практически не затрагивалась, - это построение не статических, а динамических моделей основного тона, которые позволили бы обнаруживать закономерности (индивидуальные особенности) поведения голоса в речи данного лица. До сих пор мы рассматривали только один подход построения некоторых элементов образа речевого сигнала: проецировали графики контуров частоты основного тона в пространстве (частота)*(время) на ось (частота). Было показано, что построение условных взвешенных проекций (распределений) контуров основного тона позволяет четче проявить некоторые особенности голоса, названные нами доминантными частотами основного тона голоса в речи (а в [29] названные “основными звуками”). С помощью полученного распределения оцениваются любые элементарные статистики (среднее, дисперсия, моды, коэффициент вариации и др.).

Однако вполне очевидно, что в любой указанной выше статистике отражается очень мало индивидуальных особенностей поведения голоса в речи, поскольку основная часть всех индивидуальных особенностей голоса в речи, без сомнения, содержится в динамических характеристиках контура основного тона. Поэтому, чтобы обнаружить их, необходимы динамические модели описания поведения контуров частоты основного тона. Но их пока нет. Чтобы приблизиться к их созданию, нужны новый подход, новый метод анализа динамики контуров основного тона.

Каким должен быть новый подход анализа динамики основного тона? На наш взгляд, им станет анализ траекторий контуров основного тона в фазовом пространстве. Действительно, как известно, фазовый портрет удобен для анализа прерываемых плавно меняющихся параметров, каковым и является контур основного тона. Все виды вышеупомянутых выборочных плотностей условного распределения частоты основного тона в фазовом пространстве трактуются как выделение в фазовом пространстве областей, где фазовая траектория находится чаще всего.

Для иллюстрации преимущества анализа контуров частоты основного тона в фазовом пространстве рассмотрим простейший пример аудитивного и акустического анализа вибрации голоса. Известно, что у певца, продуцирующего вибрато, контур основного тона может иметь колебания, близкие к периодическим. При этом средняя частота колебания контура составляет 6-7 Гц, а размах колебания в среднем составляет примерно 6% от среднего значения частоты основного тона [48]. Наблюдаются разные формы колебания контура частоты основного тона: близкие к гармонике, циклоиде, пере-

вернутой циклоиде и др. Рассмотрим гипотетический случай идеального гармонического вибратора, описываемого в пространстве (частота)*(время) гармонической функцией

$$F(t) = F_0 + B \cdot \sin(vt). \quad (147)$$

Фазовый портрет такого контура частоты основного тона имеет форму эллипса (см. рис. 35). Выборочная ненормированная плотность распределения частоты основного тона в этом случае (при бесконечно малом шаге параметрического описания речевого сигнала) будет иметь вид, близкий к следующему:

$$h(F) = \frac{2N}{v \cdot \sqrt{B^2 - (F - F_0)^2}}, \quad (148)$$

где N - число периодов вибратора в анализируемом сигнале.

В реальных расчетах выборочная плотность распределения, естественно, будет отличаться от теоретической (148), поскольку в них шаг параметрического описания речевого сигнала всегда конечен. Форма будет близка к (148). Это распределение - бимодальное, и средняя частота F_0 по ее доле в выборке является не то что наиболее встречаемой, а, наоборот, - наименее встречаемой из всех продуцируемых частот.

В то же время аудитор с хорошим музыкальным слухом достаточно точно определяет именно частоту F_0 , а вовсе не $F_0 - B$ или $F_0 + B$ [48]. Если применительно к данному идеализированному случаю использовать простейшую динамическую модель описания динамики основного тона в фазовом пространстве

$$\alpha \cdot F'' + \beta \cdot F' + \gamma \cdot F = \delta, \quad (149)$$

где параметры $\alpha, \beta, \gamma, \delta$ являются параметрами модели и определяются на основании имеющейся статистической выборки контура частоты основного тона, то искомые параметры F_0 и v будут иметь следующие значения:

$$v^2 = \gamma / \alpha, \quad F_0 = \delta / \gamma. \quad (150)$$

Применима ли подобная простейшая линейная модель к реальным явлениям - это самостоятельная проблема. Но то, что анализ огибающей траектории основного тона в фазовом пространстве путем построения динамических моделей более перспективен, чем тот или иной статистический анализ отдельных значений частоты основного тона, - не вызывает сомнения.

В качестве иллюстрации эффекта вибрации голоса в речи можно обратить внимание на близкую к только что описанной траектории основного тона на последнем звуке слова “мыло”, фазовый портрет которого приведен на рис. 33. На рисунке явно виден почти заверченный один цикл вибрации голоса в конце фонации (левая часть контура). Найти по этой части фазово-

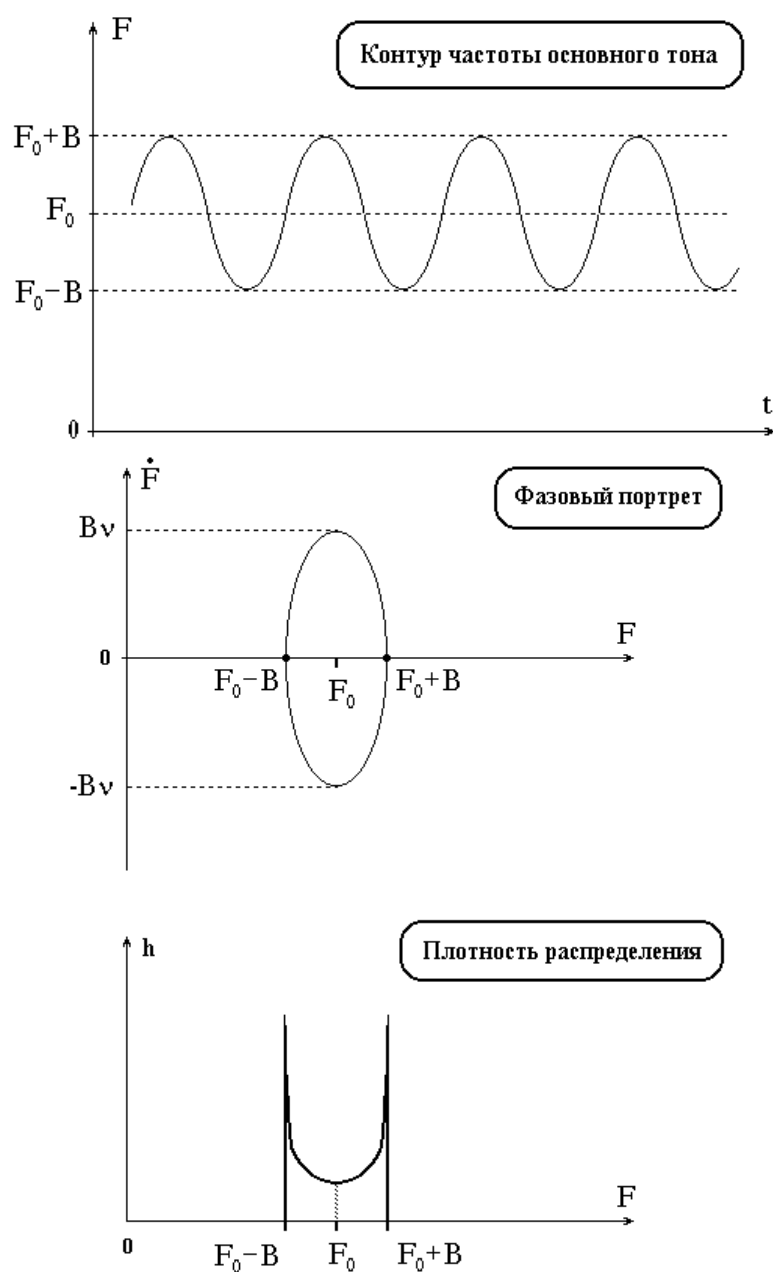


Рис. 35. Вибрато голоса

го портрета среднюю частоту, относительно которой происходит вибрация голоса, с помощью динамической модели (150) принципиально несложно. И это значение будет действительно той средней продуцируемой частотой, не искаженной вибрацией.

Перспективность построения динамических моделей основного тона голоса в речи может быть подтверждена и следующей выдержкой из работы И.И.Левидова: “Характер вибрации индивидуален у каждого голоса и в некоторой мере определяет его качество. Таким образом, вибрация может быть в известной степени рассматриваема даже как некий элемент тембра голоса. Характер вибрации определяется с детства и в неизменном виде остается на всю жизнь” [29. С.181].

К этому можно еще добавить, что изучение нового типа сонограмм - амплитудно-фазовых - позволит увидеть не только треки всех обертонов голоса, но и в цвете скачки их фазовых параметров, которые наиболее характерно будут проявляться на переходных участках между отдельными звонкими фонемами.

Влияние артикуляции на динамику голоса

В лингвистических исследованиях речи поведение голоса относится к просодическим признакам, которые несут определенную смысловую, интонационную, эмоциональную и иную полезную нагрузку. Однако постоянное и порой резкое изменение конфигурации речевого тракта, как правило, ведет к значительным изменениям контура частоты основного тона, особенно хорошо проявляющимся на переходных участках от открытых гласных звуков к звонким согласным, и наоборот. Такие относительно плавные скачки (скорость изменения частоты основного тона на них сопоставима со скоростью артикуляции) не несут какой бы то ни было нагрузки, полезной с точки зрения передачи информации слушателю. В большей мере они несут информацию об индивидуальных особенностях взаимосвязи артикуляции и голоса у данной личности.

Замечено, что у лиц с очень монотонной речью вариации голоса на переходных участках фонем существенно слабее, чем у лиц с яркой выразительной речью. Причем, говоря о выразительности и яркости речи, здесь можно различить полярные случаи, когда речь может быть объективно монотонной (на что указывают идеально ровные горизонтальные контуры интонограммы или контуры обертонов на узкополосной амплитудной сонограмме), но при этом оставаться прекрасно разборчивой (на широкополосной сонограмме четко видны сегменты каждой фонемы в отдельности). А может быть интонационно весьма подвижной (на интонограмме траектория высоты голоса живо извивается), но весьма неразборчивой (широкополосная сонограмма выглядит как некий “монолит” из одной фонемы).

Зная эти особенности речеобразования, эксперт также их учитывает в своих исследованиях индивидуальных признаков речи. Причем здесь уже нельзя четко разделить голосовые и артикуляционные признаки, поскольку в данном случае исследуются не каждый из них в отдельности, а в их взаимосвязи.

Примеры, иллюстрирующие сказанное, приведены на рис. 36, на которых показаны интонограммы тех же лиц и тех же слов, которые были приведены на рис. 27-29.

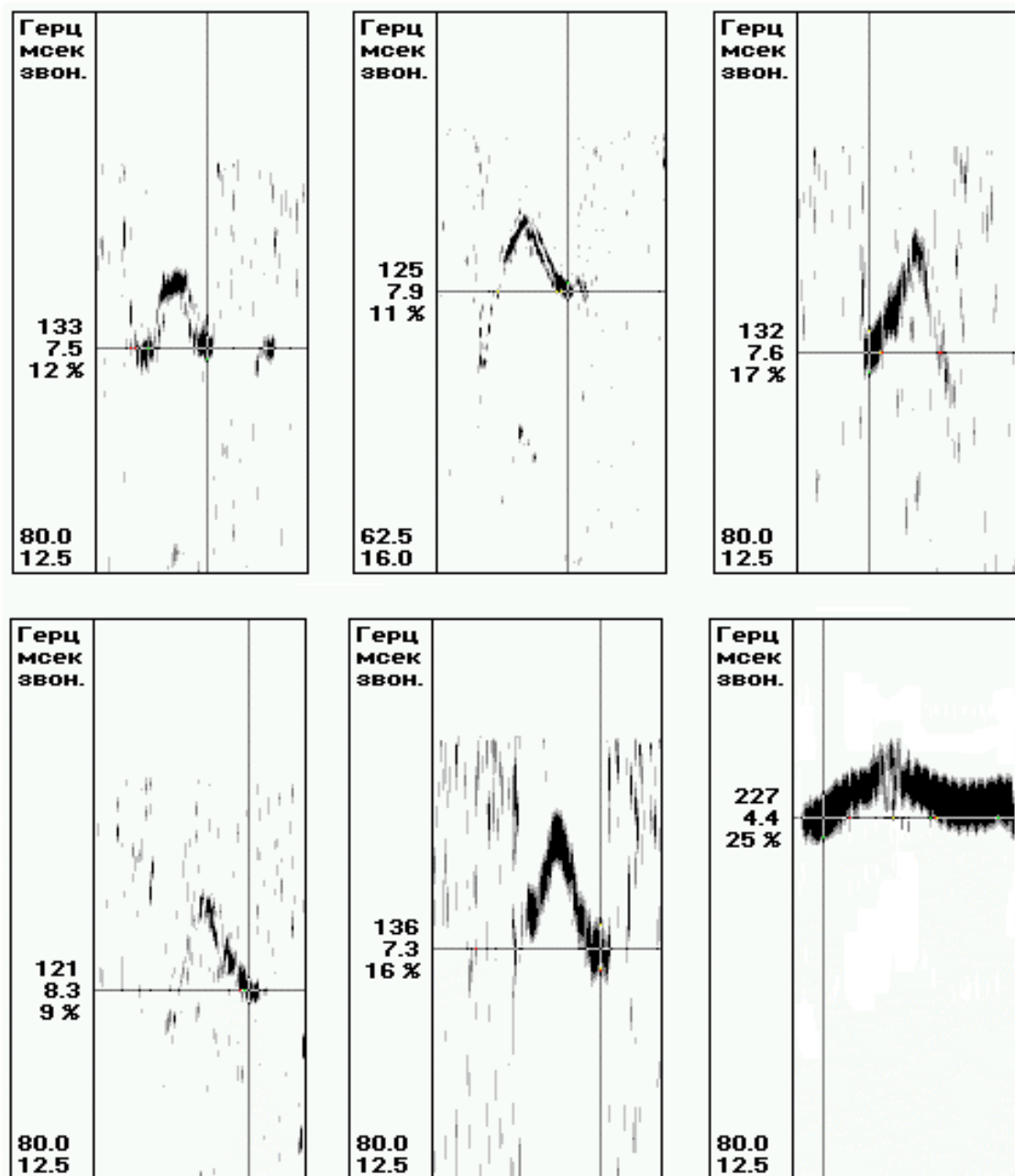


Рис. 36. Интонограммы слова “алло” разных лиц

§3. Техническое исследование фонограмм

Из всего множества вопросов технического исследования фонограмм рассмотрим только два наиболее часто встречающихся: исследование признаков монтажа фонограммы и идентификация магнитофона по его следам на магнитной ленте.

В любой экспертизе помимо основного вопроса - идентификации личности по речи - практически всегда ставится вопрос о монтаже фонограммы. Но поскольку понятие монтажа включает в себя умысел, что не входит в компетенцию эксперта-криминалиста, то он может лишь отвечать на вопрос о признаках монтажа или об аутентичности фонограммы. В действительности нарушение целостности фонограммы может быть вызвано многими причинами и не иметь какого-либо умысла. Например, может быть случайно нажата клавиша магнитофона, приводящая к его полной или временной остановке. Некоторые модели магнитофона могут сами останавливаться при отсутствии полезного сигнала и снова включаться на запись при появлении речевого сигнала. Или во время очередного прослушивания результатов записи может произойти нечаянное включение новой записи вместо включения воспроизведения, что в итоге оставляет новые следы на первоначальной записи.

Все это приводит к нарушению целостности фонограммы. Следы включения и выключения хорошо обнаруживаются не только инструментально, но и аудитивно. Более того, в экспертной практике на самом деле именно аудитивно сначала находятся все очевидные следы нарушения аутентичности фонограммы, которые затем исследуются инструментально и закрепляются документально в экспертном заключении. Однако профессионально выполненный монтаж иногда на слух не обнаруживается. В этом случае на помощь приходят физические методы исследования фонограммы.

Один из наиболее распространенных методов исследования признаков монтажа фонограммы заключается в анализе постоянства уровня шумов во всем диапазоне частот. Изменение этого уровня служит признаком нарушения аутентичности фонограммы, который следует тщательно изучить на предмет установления его происхождения.

Но в исследовании признаков монтажа наиболее эффективными оказываются такие параметры сигнала фонограммы, которые вообще не вос-

принимаются на слух. К ним, в первую очередь, относятся низкочастотные и высокочастотные составляющие гармонического типа: следы фона электросетевого питания в низкочастотной области и следы высокочастотного подмагничивания записываемого сигнала. Последнее удастся обнаружить далеко не во всех случаях, а только лишь на магнитных лентах высшего качества.

Признаки монтажа по следам фона электросетевого питания чаще всего проявляются следующим образом:

а) если на узкополосной амплитудной сонограмме обнаруживаются две фоновые линии, обычно разной интенсивности и отстоящие друг от друга на доли герца (от 0,1 до 1-5 Гц), то это указывает на то, что данная фонограмма копировалась и, может быть, к ней была “примешана” (микширована) некоторая другая фонограмма. Это должно составить предмет самостоятельного исследования;

б) если на амплитудной сонограмме обнаруживается одна фоновая линия, имеющая явный излом в форме скачка, то это является признаком нарушения скорости транспортировки ленты. Это также должно составить предмет самостоятельного исследования причин излома - либо производилась остановка записи, либо возникали какие-то нарушения в работе лентопротяжного механизма;

в) если на амплитудной сонограмме имеется одна непрерывная фоновая линия, но на амплитудно-фазовой обнаруживается скачок ее цвета, то этот случай аналогичен предыдущему.

Вопрос идентификации магнитофона по его следам на магнитной ленте решается многими способами, начиная от магнитооптических методов визуализации и исследования магнитных дорожек на ленте и кончая исследованием самих сигналов. Очень много информации для решения этого вопроса обычно содержится в стоповых импульсах стирающей магнитной головки. Для примера на рис. 37 приведены стоповые импульсы магнитных головок разного типа: в виде постоянного магнита и в виде электромагнита.

Поскольку настоящая работа не включает в себя чисто криминалистические проблемы идентификации магнитофонов по их следам на магнитной ленте, то мы лишь вкратце отметим один весьма важный аспект, ранее игнорируемый некоторыми экспертами-фоноскопистами, слабо владеющими достижениями естественных наук.

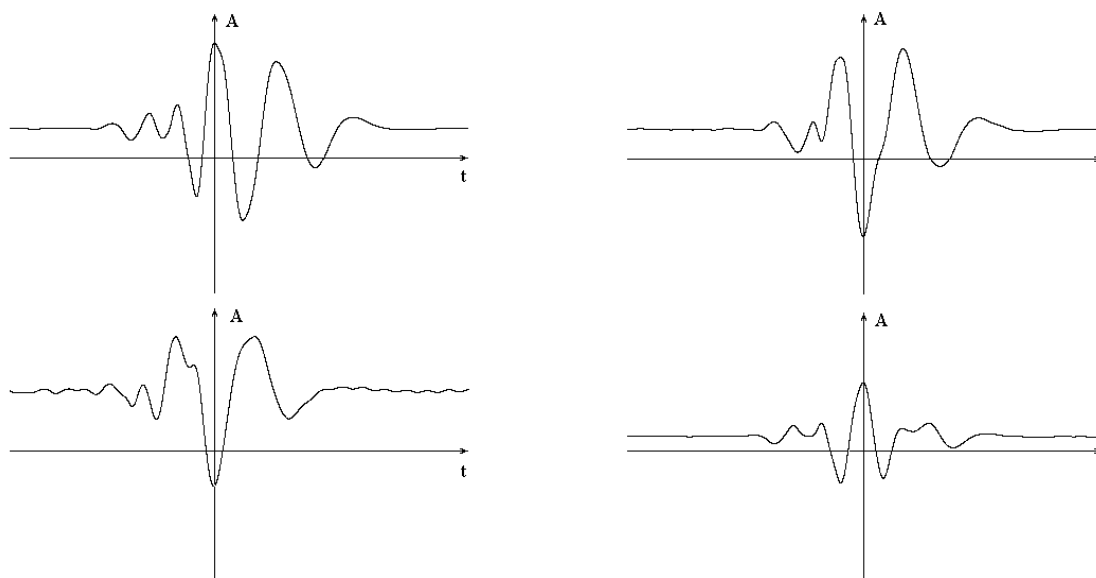


Рис. 37. Стоповые импульсы стирающих магнитных головок разного типа

Речь идет о самом понятии следа в криминалистической фоноскопии. Реально материальный след остается на магнитной ленте в виде остаточной ее намагниченности. Единственным до последнего времени прямым способом визуализации этого следа является обнаружение слабых магнитных полей с помощью специальных оптомагнитных сред. Применяя обычный микроскоп с поляризационными световыми фильтрами, можно увидеть остаточную намагниченность ленты, если ее окружить оптомагнитной средой, вращающей поляризованный свет в зависимости от направления и силы магнитного поля. В результате на поверхности магнитной ленты четко проявляются треки магнитных дорожек, оставляемых магнитными головками магнитофона. Таким образом, криминалистическая фоноскопия включает в себя помимо всех прочих ранее перечисленных направлений еще и трасологическое исследование.

Однако исследовать магнитные фонограммы в полном объеме с помощью указанных средств пока еще проблематично из-за сложности самой оптико-магнитной технологии. Проще воспроизвести магнитную фонограмму обычным способом и увидеть (пусть и значительно преобразованный в процессе воспроизведения) след, чем изучать его в “первозданном” виде, но в очень неудобной для фиксации, обработки и анализа форме. Поэтому к прямому исследованию следов магнитных треков на ленте практически никто из экспертов не прибегает. Но с развитием необходимой технологии трасологические исследования в криминалистической фоноскопии несомненно займут должное положение, поскольку существует ряд кримина-

листически важных признаков фонограммы, которые можно обнаружить исключительно лишь путем прямого изучения магнитных треков на исследуемой ленте. Однако поскольку эти проблемы выходят за рамки настоящей работы, то рассматриваться они здесь не будут.