СПОСОБ ФОРМИРОВАНИЯ ФОНОГРАММ РЕЧЕВОЙ ИНФОРМАЦИИ

Д.т.н., профессор В.Р.Женило, к.т.н., доцент В.И.Кирин (Академия управления МВД России), М.В.Женило (Московский технический университет связи и информатики), С.В.Женило (Московский физико-технический институт)

В докладе излагается запатентованный «Способ формирования фонограмм речевой информации», обеспечивающий защиту фонодокументов от их фальсификаций.

Уточним само понятие «фонодокумента», используемое в данном докладе. Фонодокумент – это стандартный компьютерный звуковой файл, полученный следующим образом:

а) в память компьютера поступает очередная порция (блок) оцифрованного полезного электроакустического сигнала, с того или иного входа цифрового регистратора речи;

б) параллельно с этим блоком создается блок сгенерированного защитного сигнала со следующей структурой – это набор из N частотно модулированных гармоник[1] постоянной амплитуды с эпизодически меняющейся случайным образом начальной фазой;

в) в зависимости от динамики мощности частотно-временных компонентов полезного сигнала к нему прибавляется сгенерированный защитный сигнал, частотные компоненты которого специальным образом амплитудно модулируются частотными компонентами полезного сигнала, поступившего на вход системы регистрации;

г) защищенный таким образом в оперативной памяти компьютера «на лету» очередной блок фрагмента будущей цифровой фонограммы записывается в файл, который далее будет называться цифровым «фонодокументом».

Основываясь на перечисленных требованиях и понятии фонодокумента, можно предложить следующий способ его формирования. Но, чтобы не вдаваться в технические подробности описания подмешивания в документируемый полезный сигнал защитной сигнальной структуры «налету в оперативной памяти компьютера», будем считать, что для защиты фонограммы она сначала запоминается в полном объеме на некотором носителе. Параллельно с этой цифровой фонограммой формируется цифровой защитный сигнал, состоящий из набора частотно модулированных гармоник постоянной амплитуды, частотный диапазон которых перекрывает весь частотный диапазон защищаемого речевого сигнала, и которые сами с собой никогда не пересекаются по частоте. Структура частотной модуляции защитных сигналов выбирается такой, чтобы они никогда не повторялась и содержала информацию о времени, типе и номере устройства звукозаписи. Начальная амплитуда всех защитных частотно модулированных гармоник устанавливается равной максимально допустимой амплитуде помех, принятой в данной технологии записи фонограммы. Затем на разных частотно-временных участках защитный сигнал модулируется по амплитуде синхронно защищаемому речевому сигналу следующим образом.

Весь речевой сигнал нарезается на отдельные кадры, составляющие сонофильм. Каждый кадр сонофильма имеет свою спектральную мощность. Синхронно с этими кадрами сонофильма речевого сигнала создаются кадры фильма защитного сигнала. Спектральная мощность кадров защитного сигнала устанавливается за два прохода анализа кадров речевого сигнала в прямом и обратном направлении времени смены кадров.

При прямом проходе по кадрам фильма логарифмический амплитудный спектр защитного сигнала $P(c,f)$ в полосе частот f в кадре c устанавливается следующим образом:

$$P(c,f) = max \{ N(f); S(c-1, f) – M(f) – D^{+}(f); G(c,f) \},$$

где $N(f)$ – максимально допустимый уровень помех в полосе частот f в фиксируемой фонограмме;
$S(c, f)$ – логарифм амплитудного спектра речевого сигнала в кадре c в полосе частот f;
$M(f)$ – функция, задающая разницу самого мощного участка речевого сигнала и логарифма амплитудного спектра защитного сигнала в ближайшей частотно-временной окрестности;
$D^{+}(f)$ – функция, задающая на сколько ниже должен быть логарифмический амплитудный спектр защитного сигнала в следующем кадре, чтобы эффект психоакустической маскировки сигналов сработал в полной мере;
$G(c,f)$ – начальный уровень мощности защитного сигнала во всех кадрах c и каналах частот f.

При обратном проходе по кадрам фильма логарифм амплитудного спектра защитного сигнала $P(c,f)$ в полосе частот f в кадре c устанавливается следующим образом:

$$P(c,f) = max \{ N(f); S(c+1, f) – M(f) – D^{-}(f); P(c,f) \},$$

где $D^{-}(f)$ – функция, задающая на сколько ниже должен быть логарифмический амплитудный защитного сигнала в следующем кадре, чтобы эффект психоакустической маскировки сигналов сработал в полной мере.

Защитный сигнал – это суперпозиция частотно модулированных гармоник с частотами, не выходящими за границы треть октавной полосы, и с эпизодически случайно меняющимися начальными фазами. Динамика амплитуды защитного сигнала меняется в зависимости от изменения динамики уровня мощности речевого сигнала в полосе частот f. Причем, динамика амплитуды защитного сигнала на максимальных по мощности кадрах речевого сигнала в полосе частот f должна быть ниже последнего, например, на –20 дБ (по требованию системы

[1] Средние значения частот этих гармоник, их частота и начальная фаза частотной модуляции могут нести информацию об идентификационном номере регистратора, времени записи и т.п.

идентификации лиц по фонограммам русской речи «Диалект»[1]). А после максимального участка амплитуда защитного сигнала экспоненциально снижается. Это свойство приводит к тому, что из-за эффекта маскировки защитный сигнал практически не слышен, но его уровень мощности часто оказывается выше текущего уровня мощности регистрируемого речевого сигнала (особенно часто это возникает на участках звуковых пауз в речевом сигнале в разных частотных диапазонах). Благодаря этому свойству защитную структуру сигнала можно визуализировать с помощью обычных динамических спектральных фильмов (сонограмм, сонофильмов, фасфильмов).

В силу того, что на самых мощных частотно-временных участках полезного (речевого) сигнала защитный сигнала всегда ниже по амплитуде на -20 дБ, то защищенный речевой сигнал остается пригодным для проведения экспертно-криминалистических идентификации личности по следам речевого сигнала.

А из-за малой мощности сигналов защитной структуры частота, амплитуда и фаза которых постоянно меняется, их практически невозможно вычислить и бесследно удалить.

Если в фонограмме, защищенной таким образом, попытаться бесследно выделить и удалить некоторый фрагмент речи, то вместе с этим фрагментом речевого сигнала удаляется и защитная структура, нарушение которой проявляется и обнаруживается на сонограммах в силу высокой информационной избыточности и неповторяемости взаимного расположения элементов защитной структуры.

Если попытаться внести в защищенную цифровую фонограмму фрагмент незащищенного речевого сигнала, то на участке монтажа защитная структура окажется не модулированной этим речевым сигналом, что тоже обнаруживается с помощью сонограммы, и свидетельствует о нарушении аутентичности цифровой фонограммы.

Если попытаться переставить местами фрагменты защищенной фонограммы с наложением или без ее отдельных частей, то произойдет такая же перестановка и наложение и защитной структуры, что также обнаруживается с помощью сонограммы.

Структура защитного сигнала не случайно выбрана в виде частотно модулированных гармоник с некратными частотами модуляций и случайно меняющимися фазами. С одной стороны, такие гармоники достаточно регулярны для того, чтобы их можно было обнаружить спектральными методами. Но, с другой стороны, начальные фазовые характеристики этих гармоник меняется случайным образом так, чтобы их нельзя было точно вычислить для удаления или подмены. Защитный сигнал по амплитуде модулируется уровнем мощности речевого сигнала в соответствующих полосах частот таким образом, чтобы, во-первых, благодаря психоакустическому эффекту маскировки звуков не быть заметным. Во-вторых, быть уникальным в привязке к меняющейся спектральной мощности полезного речевого сигнала. И, в-третьих, быть достаточно слабым для возможности использования защищенного речевого сигнала в экспертно-криминалистической идентификации личности по речи.

Частотно модулированные гармоники защитного сигнала выбираются таким образом, чтобы они перекрывали весь частотный диапазон речевого сигнала и их частоты модуляции не были взаимно кратными. Начиная с $N = 2$, при условии не кратности частот модуляции гармоник защитного сигнала любая вырезка внутренней части фрагмента фонограммы может проявиться в разрыве динамики частот гармоник защитного сигнала. Однако, из-за слабости гармоник защитного сигнала иногда двух гармоник оказывается мало. Поэтому, чтобы гарантировать большую защищенность фонограммы, число гармоник защитного сигнала следует увеличивать. В настоящее время в компьютерном регистраторе оперативных служебных сообщений (в системе «КРОСС-документ») таких гармоник восемь.

Для иллюстрации результатов применения предложенной технологии на рис. 1 показан сонофильм некоторого исходного синтезированного звукового файла, первоначально незащищенного никакими дополнительными сигнальными структурами. Динамический диапазон визуализируемых следов этого синтезированного сигнала на рис. 1 выбран от 0 до –60 дБ. На рис. 2–6 приведены пять сонофильмов того же синтезированного сигнала, но уже защищенного дополнительной сигнальной структурой. Динамический диапазон визуализации следов на этих сонофильмах составляет от 0 до –20 дБ (на рис. 2) и от 0 до –60 дБ (на рис.6).

[1] Патент на изобретение «Способ идентификации личности по фонограммам произвольной устной речи» - RU – 2107950 – C1 – 6G10L 5/06.

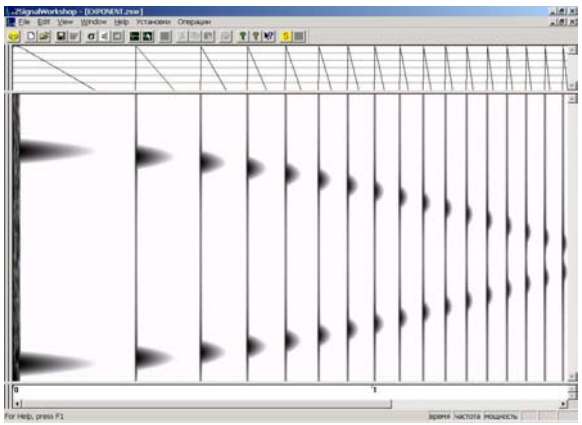

Рис. 1. Уровень мощности (вверху) исходного незащищенного синтетического сигнала и его сонофильм (внизу) с диапазоном визуализации амплитудных спектров от 0 до –60 дБ.

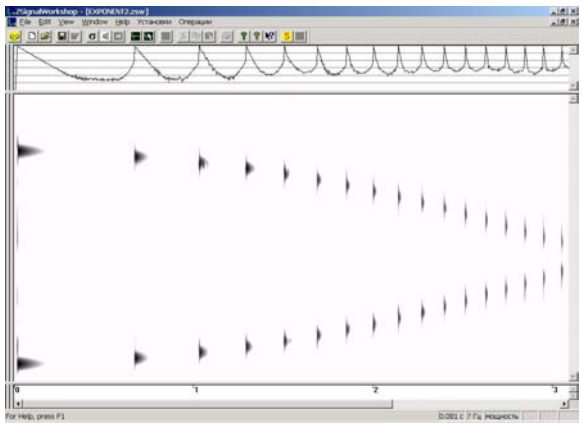

Рис. 2. Уровень мощности (вверху) защищенного сигнала с рис. 1 и его сонофильм (внизу) с диапазоном визуализации амплитудных спектров от 0 до –20 дБ.

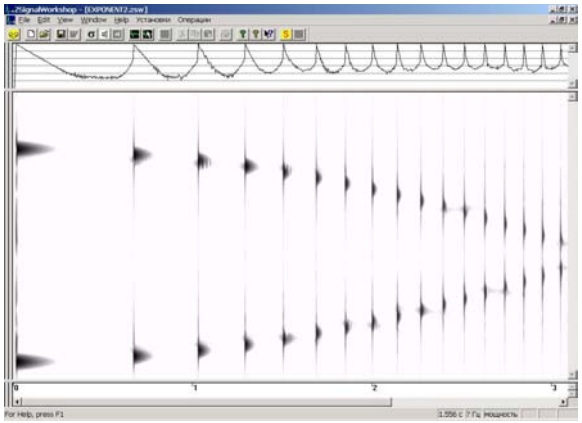

Рис. 3. Уровень мощности (вверху) защищенного сигнала с рис. 1 и его сонофильм (внизу) с диапазоном визуализации амплитудных спектров от 0 до –30 дБ.

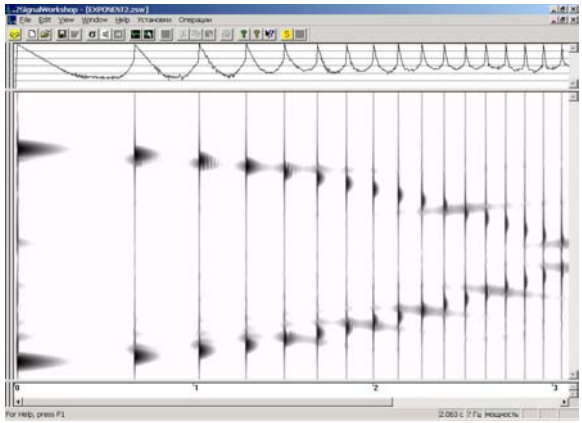

Рис. 4. Уровень мощности (вверху) защищенного сигнала с рис. 1 и его сонофильм (внизу) с диапазоном визуализации амплитудных спектров от 0 до –40 дБ.

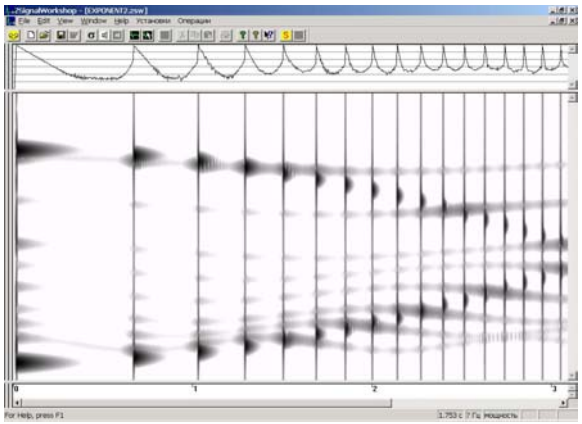

Рис. 5. Уровень мощности (вверху) защищенного сигнала с рис. 1 и его сонофильм (внизу) с диапазоном визуализации амплитудных спектров от 0 до –50 дБ.

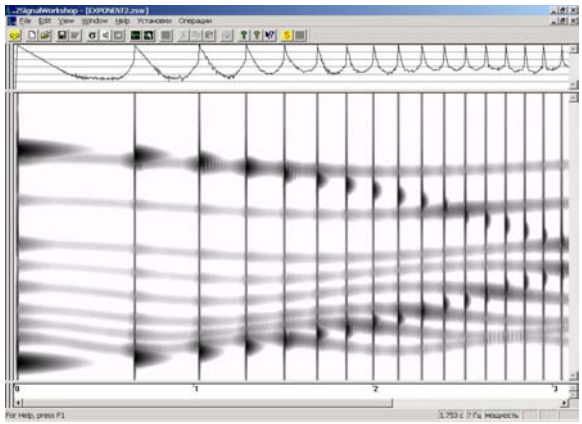

Рис. 6. Уровень мощности (вверху) защищенного сигнала с рис. 1 и его сонофильм (внизу) с диапазоном визуализации амплитудных спектров от 0 до –60 дБ.

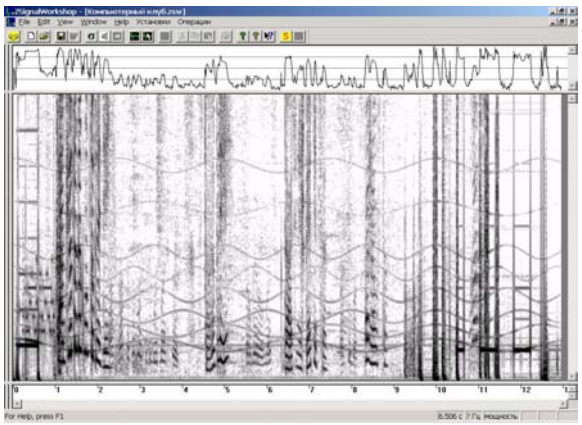

Рис. 7. Уровень мощности защищенного речевого сигнала телефонного качества и его сонофильм.

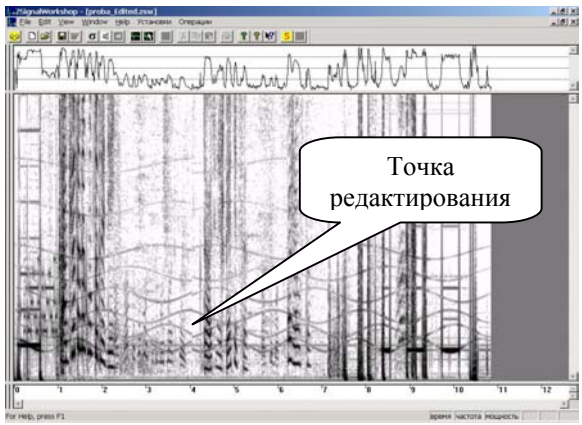

Рис. 8. То же, что и на рис. 7, но речевой сигнал отредактирован.

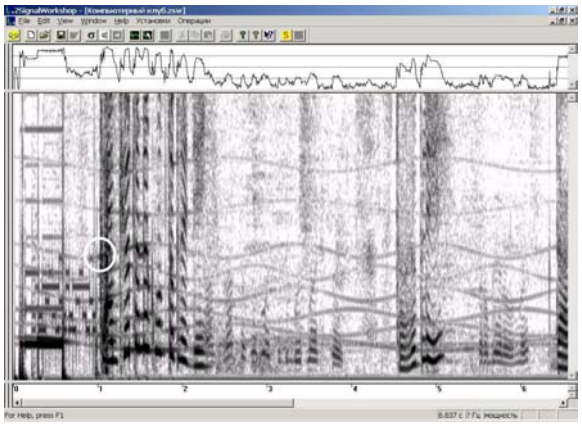

Рис. 9. Сонофильм со спектральным разрешением 24 Гц (внизу) части следа защитного сигнала с рис. 7 (на рисунке отмечено белым кружком).

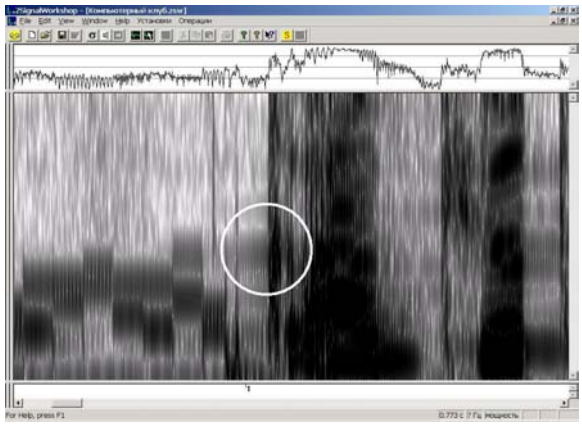

Рис. 10. Сонофильм со спектральным разрешением 192 Гц (внизу) части следа защитного сигнала с рис. 7 (на рисунке отмечено белым кружком).

Как визуализируются гармоники защитного сигнала на реальных речевых сигналах (телефонного качества), показано на рис. 7, где приведен частотно-временной образ фонодокумента, полученного с помощью системы «КРОСС-документ». Если у такого фонодокумента попытаться удалить какой-либо фрагмент, то следы от этой манипуляции проявятся в нарушении структуры защитного сигнала. В качестве примера на рис. 8 показан результат монтажа этого фонодокумента - в нем на участках речевых пауз вырезан небольшой фрагмент с 4-ой по 6-тую секунду. Следует отметить, что динамика спектра шумов на выбранных участках монтажа такова, что, если бы не было защитных сигналов, то обнаружить нарушение однородности этих шумов было бы просто невозможно. По этой причине *бесследный* перенос и наложение любых фрагментов внутри одного и того же фонодокумента оказывается практически невозможным.

Невозможным оказывается и добавление к имеющемуся фонодокументу чистых речевых сигналов из других цифровых фонограмм. Дело в том, что защитные сигналы структурно привязываются к полезному первоначальному сигналу. И если эта привязка не произошла, то этот факт обнаруживается при более тщательном спектрально временном анализе фонодокумента. Для пояснения сказанного на рис. 9-10 приведены сонофильмы одного и того же фрагмента фонодокумента, но эти оптимальные сонофильмы имеют разные частотно-временные разрешения.

На всех рисунках виден след одной из частотно модулированной гармоники защитной структуры, обведенный белым кружком, который указывает на одну и ту же частотно-временную точку сонофильма исследуемого фонодокумента, на которую следует обращать внимание при решении вопроса о присутствии следа речевого сигнала в момент рождения фонодокумента.

Признаком того, что на рис. 8 имеется не примешанный позже чистый речевой сигнал, полученный не в момент создания фонодокумента, является наличие следа одной из гармоник защитного сигнала. Причем, эта гармоника амплитудно-модулирована следом полезного речевого сигнала (в данном случае амплитуда защитного сигнала растет перед следом полезного сигнала).

Предложенная технология защиты звуковых (речевых) сигналов реализуема не только на базе вычислительной техники. Она может быть выполнена в виде отдельного устройства и на аналоговой аппаратуре (правда в усеченном виде, поскольку реализовать обработку аналогового сигнала в обратном порядке следования сигнала довольно сложно). Приведем краткое его описание.

В устройство информационной защиты речевого сигнала поступает исходный оригинальный речевой сигнал. Без каких-либо искажений он поступает на сумматор защитных сигналов и выходит из устройства. Формирование структуры защитных сигналов управляется самим речевым сигналом. Структура защитных сиг-

налов формируется с помощью нескольких однотипных блоков следующим образом. Оригинальный исходный речевой сигнал поступает на гребенку треть октавных полосовых фильтров, перекрывающих диапазон частот от 100 до 4000 Гц. Амплитуда выходного сигнала каждого из этих фильтров уменьшается 20 дБ и сравнивается с пороговым значением максимально допустимого для данного класса записывающего устройства амплитуды гармонической помехи. Большее значение этих амплитуд задает амплитуду генерируемой частотно модулированной гармоники, которая является составной частью защитного сигнала. Частота этой гармоники постоянно меняется в пределах полосы пропускания соответствующего треть октавного фильтра, а ее начальная фаза случайно (например, в среднем один раз в секунду) зануляется. Получаемые таким образом в разных полосах частот защитные частотно модулированные гармонии со случайной фазой отражают структуру оригинального речевого сигнала. Все они вместе с оригинальным речевым сигналом поступают на сумматор сигналов, на выходе которого образуется исходный речевой сигнал с аддитивной структурой слабых частотно и амплитудно-модулированных гармоник.

На рис. 11 представлена общее схематическое описание устройства защиты речевых сигналов.

Общая схема устройства содержит следующие блоки: 1, 2 и 3 – устройства формирования защитного сигнала в треть октавной полосе частот (всего таких устройств – 17), 4 – сумматор сигналов. Эта схема работает следующим образом.

Исходный (входной) речевой сигнал поступает на входы 17-ти однотипных блоков формирования защитных сигналов. Каждый блок в треть октавных полосах частот, перекрывающих общий диапазон частот от 100 до 4000 Гц, формирует защитный сигнал. Защитный сигнал – это частотно и амплитудно-модулированная гармоника с частотой, не выходящей за границы треть октавной полосы, и со случайно меняющейся начальной фазой (в среднем один раз в секунду). Динамика амплитуды защитного сигнала меняется в зависимости от изменения динамики уровня мощности речевого сигнала в данной треть октавной полосе частот.

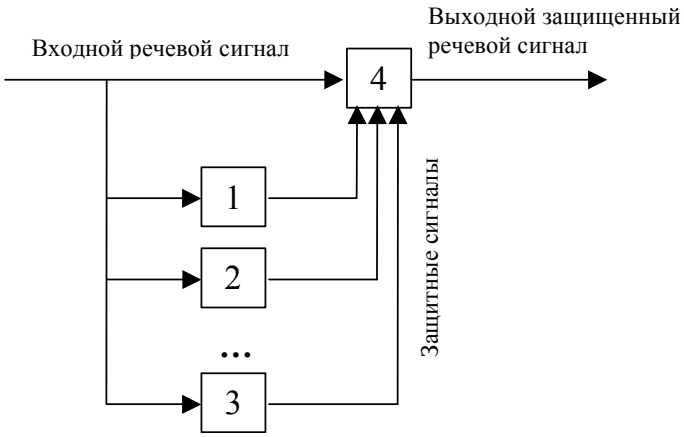

Рис. 11. Общая схема устройства защиты речевого сигнала.

На рис. 12. представлена схема устройства формирования защитного сигнала в одной треть октавной полосе частот.

Устройство содержит следующие блоки: 5 – треть октавный полосовой фильтр частот, 6 – атеньюатор сигнала на –20 дБ, 7 - детектор амплитуды сигнала, 8 – интегрирующая цепочка с заданным декрементом затухания, 9 – амплитудный селектор,10 – амплитудный селектор, 11 - генератор качающей частоты, 12 – установка порогового значения амплитуды, 13 – установка начальной фазы, 14 – генератор случайных импульсов.

Общая схема формирования защитного сигнала работает следующим образом.

Исходный речевой сигнал поступает на вход блока 5 - треть октавного полосового фильтра частот. Полученный узкополосный сигнал ослабляется в блоке 6 на – 20 дБ. Детектор сигнала в блоке 7 выделяет из поступающего на его вход амплитудную огибающую сигнала. Полученный сигнал, отражающий динамику изменения амплитуды речевого сигнала в данной полосе частот треть октавного фильтра, поступает на вход интегрирующего блок 8, который обеспечивает плавное затухание сигнала с декрементом затухания 10 мс. В блоке 9 амплитудного селектора происходит сравнение уровней сигналов с выхода блоков 7 и 8 и на выход блока 9 пропускается максимальный из них. В блоке 12 задается начальный уровень будущего защитного сигнала в соответствии с требованиями уровня гармонических помех, максимально допустимых в системе звукозаписи (для телефонии этот уровень равен –35 дБ, для магнитофонов - 48 дБ и т.п.). В блоке 10 происходит сравнение уровней сигналов с выхода блоков 9 и 12 и на выход блока 10 пропускается максимальный из них. Генератор качающей частоты 11 формирует сигнал с амплитудой, задаваемой выходом блока 10. Блок 13 в случайный момент времени (в среднем один раз в секунду) сбрасывает начальную фазу генератора качающей частоты 11. Блок 14 – это генератор случайных импульсов со средним временем появления импульса – 1 секунда

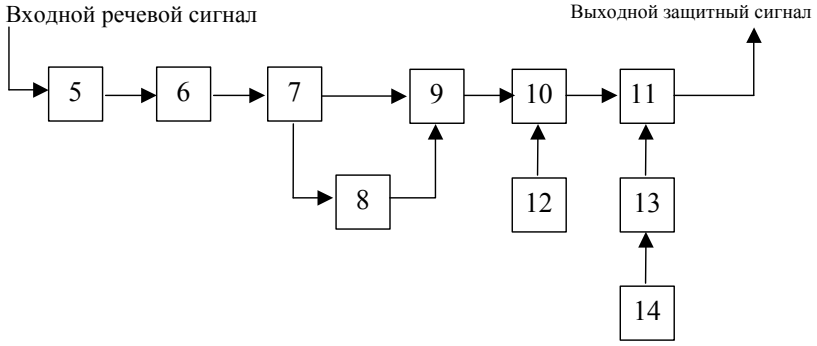

Рис. 12. Схема устройства формирования одного защитного сигнала в одной треть октавной полосе частот.

и стандартным отклонением – 0.1 с. Выход с блока 11 – это и есть защитный сигнал.

Защитный сигнал – это частотно модулированная гармоника с частотой, не выходящей за границы треть октавной полосы, и со случайно меняющейся начальной фазой (в среднем один раз в секунду). Динамика амплитуды защитного сигнала меняется в зависимости от изменения динамики уровня мощности речевого сигнала в данной треть октавной полосе частот. Причем, динамика амплитуды защитного сигнала на пиковых по мощности участках речевого сигнала в треть октавной полосе частот ниже последнего на –20 дБ. А после пикового участка амплитуда защитного сигнала экспоненциально снижается со скоростью 6 дБ за 10 мс. Это свойство приводит к тому, что из-за эффекта маскировки защитный сигнал практически не слышен, но его уровень часто оказывается выше текущего уровня регистрируемого речевого сигнала (на участках звуковых пауз в речевом сигнале в разных частотных диапазонах). Благодаря этому защитную структуру сигнала можно визуализировать с помощью динамических спектральных фильмов или сонограмм.

В силу того, что на самых мощных частотно-временных участках полезного (речевого) сигнала защитный сигнала всегда ниже по амплитуде на -20 дБ, то защищенный речевой сигнал остается пригодным для проведения экспертно-криминалистических идентификации личности по следам речевого сигнала. По требованию официально применяемой в России методике идентификации личности по речевому сигнала «Диалект» уровень помех не должен превышать –15 дБ следов речевого сигнала во всех диапазонах частот.

Из-за малой мощности сигналов защитной структуры частота, амплитуда и фаза которых постоянно меняется, их практически невозможно вычислить и бесследно удалить.

Если в фонограмме, защищенной таким образом, попытаться бесследно выделить и удалить некоторый фрагмент речи, то вместе с этим фрагментом речевого сигнала удаляется и защитная структура, нарушение которой проявляется и обнаруживается на сонограммах в силу высокой информационной избыточности, неповторяемости взаимного расположения элементов защитной структуры.

Если попытаться внести в защищенную фонограмму структуру незащищенного речевого сигнала, то на участке монтажа защитная структура окажется не модулирована этим речевым сигналом, что тоже обнаруживается с помощью сонограммы.

Если попытаться переставить местами фрагменты защищенной фонограммы с наложением или без ее отдельных частей, то произойдет такая же перестановка и наложение и защитной структуры, что также обнаруживается с помощью сонограммы.

Чем эта технология лучше других, известных в настоящее время? Тем, что подмешанный структурированный защитный сигнал, во-первых, практически не мешает восприятию речи. Во-вторых, в силу его слабости, следы его практически невозможно удалить, поскольку этому мешают мощные компоненты самого речевого сигнала. В-третьих, следы защитного сигнала постоянно меняются по частоте и амплитуде синхронно полезному (речевому) сигнала, оставаясь видными на сонофильмах. В-четвертых, структура защитного сигнала равномерно заполняет всю битовую структуру цифровой звукозаписи, неизменно следуя за изменяющимися по мощности отдельными частотными компонентами речевого сигнала. Из-за последнего невозможно бесследное (в информационном смысле структуры цифровой записи) вырывание фрагментов сигнала или вставка в сигнал иных речевых сигналов даже из других незащищенных звуковых записей.

Предложенная технология, по мнению авторов, надежно защитит цифровые звукозаписи и позволит перевести их в разряд документов в юридическом смысле. Однако следует подчеркнуть, что и в этом случае решение вопроса об отнесении той или иной конкретной фонограммы к категории документа должен решать специалист, являющийся полноправным участником того или иного уголовного процесса.

Описанная технология создания фонодокументов может использоваться не только в компьютерных многоканальных регистраторах речи, устанавливаемых в дежурных частях правоохранительных органов, службах безопасности, скорой помощи и т.п., но и в системе телефонии, радиосвязи, радио и телевещания. То есть, везде, где та или иная звукозапись может перейти из разряда простой фонограммы в разряд фонодокумента.