

Распределённая система фоноучёта «VoiceNet ID»

Тимофеев А.В.,
доктор технических наук

Введение

Опыт применения систем автоматизации фоноучёта и экспресс-исследований фонограмм речи серии «Трал» позволили специалистам ЦРТ сформулировать требования к распределённой системе фоноучёта «VoiceNet ID».

Основное назначение системы «VoiceNet ID» — хранение и оперативный поиск фонограмм, содержащих речь интересующего лица (или группы лиц).

Ключевым отличием новой системы от существующих комплексов является уникальная методология принятия обобщённого решения на основе результатов работы трёх независимых методов, а также оригинальная, трёхзвенная, архитектура программно-аппаратного комплекса (с тонким клиентом), по сути дела представляющего собой универсальное хранилище медиаданных.

Описание работы и функциональные возможности системы

Основными задачами, которые ставятся перед распределённой системой фоноучёта, являются:

- ведение федерального фоноскопического учёта для государственных органов, производящих расследование по уголовным делам;
- проведение фоноскопических исследований в интересах оперативно-разыскных подразделений;
- осуществление консультационной и учебно-методической деятельности.

В соответствии с требованиями распределённого фоноучёта система «VoiceNet ID» обеспечивает централизованное хранение и обработку информации. Ввод информации, формирование запросов на поиск (идентификацию) и проверку (верификацию) осуществляется с большого количества удалённых мест, при этом собственно доступ к данным осуществляется через Web-интерфейс. Таким образом, требования к оборудованию терминальных точек доступа к системе «VoiceNet ID» минимальны.

Система представляет собой распределённый вычислительный комплекс, включающий в себя вычислительное ядро и удалённые клиентские места. Архитектура системы обеспечивает функционирование клиентских мест по каналам Интернет с использованием Web-интерфейса. Система использует программный протокол, обеспечивающий возможность интеграции с другими информационными системами для



обмена информацией. Ядро системы функционирует под управлением операционной системы семейства UNIX. Безопасность передачи данных по открытым сетям обеспечивается протоколом HTTPS и авторизацией пользователей.

Состав системы и назначение её отдельных компонентов приводятся в таблице 1.

Таблица 1

Название компонента системы	Назначение
Брандмауэр	Обеспечение сетевой безопасности внутренней инфраструктуры вычислительного ядра
Web-сервер	Обеспечение инфраструктуры интерфейса пользователя
Диспетчер задач	Обеспечение диспетчерских функций системы
Хранилище данных	Обеспечение хранения данных и доступа к ним по запросу
Вычислительный кластер	Обеспечение параллельной отработки заданий пользователей
APM администратора	Автоматизированное рабочее место администратора
Фонолаборатория	Один из типов клиентов — криминалистическая лаборатория
Web-клиент	Наиболее массовый тип клиента — регистрация новых лиц в хранилище ядра системы, а также формирование запросов на поиск и проверку

Принцип работы системы основан на выделении из фонограмм речи и последующем попарном сравнении биометрических признаков (содержащихся в голосе индивидуальных, идентификационно значимых, признаков личности).

Выделение и сравнение индивидуальных признаков производится с использованием трёх различных языко- и текстонезависимых методов. В качестве основного используется спектрально-формантный метод, в качестве вспомогательных — метод статистик основного тона и метод на основе СГР (смеси Гауссовых распределений).

Необходимость использования одновременно трёх независимых методов обусловлена ограниченной областью применения каждого метода в отдельности, что иллюстрирует таблица 2 (количество знаков «+» отражает степень зависимости метода от параметров сигнала).

Таблица 2

Метод	Параметры сигнала		
	Продолжительность	Качество сигнала	Физическое и эмоциональное состояние
Спектрально-формантный	+++	++	+
СОТ	++	+	++++
СГР	+++	++++	++

Используемые методы поиска и параметры их надёжности

В системе «VoiceNet ID» используются языко- и текстонезависимые технологии поиска по голосу. Другими словами, неважно, что и на каком языке говорит человек. Исключение составляют т.н. тональные языки (вьетнамский, китайский, японский и т.п.), которые требуют перенастройки алгоритмов идентификации.

1. Спектрально-формантный метод

- 1.1. Данный метод основан на тезисе об уникальности геометрии речевого тракта у каждого человека и отражении данного факта в различных спектральных характеристиках речи. Наиболее явно различие спектральных характеристик проявляется в частотной ориентации и взаимном расположении формант.
- 1.2. Используемый в системе спектрально-формантный метод основан на выделении и сравнении положения и динамики поведения трёх и более формант. Данный метод защищён российским патентом.
- 1.3. Применение спектрально-формантного метода обеспечивает значение EER~8%. Значение EER для конкретного случая зависит от длительности и качества сравниваемых речевых фрагментов.
- 1.4. Данный метод является основным по следующим причинам:
 - 1.4.1. Метод предъявляет самые низкие, по сравнению с другими, требования к качеству сигнала. Возможна работа с сигналами вплоть до отношения сигнал/шум 12 дБ.
 - 1.4.2. Метод демонстрирует сравнительно высокую скорость выделения биометрических признаков и относительно робастен к типу канала.

2. Метод статистик основного тона

- 2.1. Данный метод использует шестнадцать различных характеристик основного тона (ОТ) голоса: среднее значение ОТ, максимальное значение, минимальное значение, медиана, процент участков с возрастающим тоном, дисперсия логарифма тона, асимметрия логарифма тона, эксцесс логарифма тона и другие параметры.
- 2.2. Значение EER для метода статистик основного тона зависит от длительности сравниваемых речевых фрагментов и может достигать величины ~16%. Необходимо отметить, что реализация данного алгоритма стала возможной благодаря созданию специалистами ЦРТ полностью автоматического высокоточного выделителя основного тона.
- 2.3. Достоинством данного метода является высокая скорость сравнения признаков и, как следствие, высокая скорость поиска или проверки личности. В то же время зависимость надёжности данного метода от эмоционально-психологического состояния диктора в момент произнесения позволяют использовать его в «VoiceNet ID» лишь в качестве вспомогательного.

3. Метод на основе СГР

- 3.1. Модели Гауссовых смесей (Gaussian Mixture Models) или смеси Гауссовых распределений на данный момент являются наиболее распространённым подходом к решению задач текстонезависимой идентификации.
- 3.2. Суть метода состоит в моделировании дикторозависимых акустических особенностей в пределах индивидуальных фонетических звуков (классов), которые входят в состав речевого сигнала. Сравняя дикторозависимые акустические особенности в произ-



несении одного диктора с акустическими особенностями произнесения другого диктора, можно получить меру отличия дикторов в пространстве признаков.

- 3.3. Значения EER для метода на основе СГР зависят от длительности сравниваемых речевых фрагментов и могут достигать величины ~4–5%.
- 3.4. Высокая требовательность метода к качеству сигнала, высокая зависимость от обучающего материала, а также относительно большие временные затраты на выделение биометрических признаков не позволяют использовать его в системе в качестве основного. Впрочем, использование связи SVM-GMM позволило решить задачу каналокомпенсации путём понижения показателя EER на 2–3% по сравнению со связкой NN-GMM, используемой традиционно.

Аппаратные характеристики

Для обеспечения максимальной производительности системы целесообразно использовать блейд-серверы IBM на 19-дюймовом шасси. Сервер хранения базы данных рассчитывается с учётом объёма, занимаемого учётными карточками, и требований обеспечения быстродействия при обработке запросов. На хранение одной учётной карточки в базе данных отводится 7,5 МБ, из которых:

- 5,5 МБ — на хранение звукового сигнала, средняя продолжительность которого 4 мин. (включая паузы; в формате ИКМ 16 бит при частоте дискретизации 11025 Гц);
- 500 кБ — на хранение биометрической информации;
- 500 кБ — на хранение дополнительной информации о личности (установочные данные, информация из других информационных систем);
- 1 МБ — резерв.

Для реализации в полном объёме требуемых характеристик системы (что на практике означает проведение свыше 700 млн. сравнений в сутки) может потребоваться до 60 восьмиядерных блейд-серверов и сервер базы данных на 15 ТБ. Энергопотребление комплекса (с резервом) составит 100 кВт. Все серверы оснащаются блоками бесперебойного питания. С учётом постоянного повышения производительности компьютеров и совершенствования алгоритмов идентификации заявленное количество серверов может корректироваться.

В высокой степени надёжность, безопасность, высокая производительность комплекса обеспечиваются применением СУБД Oracle. В частности, СУБД Oracle позволяет работать с информацией практически неограниченному числу пользователей (при наличии достаточных аппаратных ресурсов), не проявляя тенденции к снижению производительности системы при резком увеличении их числа.

Механизмы масштабирования СУБД Oracle последней версии позволяют практически безгранично увеличивать мощность и скорость работы сервера базы данных и приложений простым добавлением новых узлов (серверов) кластера. Это не требует остановки и модернизации уже работающих приложений. Кроме того, выход из строя отдельных узлов кластера также не приводит к остановке приложения.

Тимофеев А.В.

Распределённая система фоноучёта «VoiceNet ID»

Для защиты кластера от перебоев электропитания предусмотрено его оснащение системой непрерывного питания.

Суточная загрузка системы представлена в таблице 3.

Таблица 3

Суточная нагрузка «VoiceNet ID»

Число вычислительных «лезвий»	8	18	38	58
Число ядер на «лезвии»	16	16	16	16
Общее число ядер	128	288	608	928
Максимальная расчётная загрузка кластера в сутки (количество попарных сравнений)	27 648 000	62 208 000	160 000 000	600 000 000

Заключение

Разработанные ЦРТ передовые средства и методы обработки речевых сигналов позволяют уже сегодня автоматизировать, с использованием распределённой системы фоноучёта, большую часть операций, связанных с поиском диктора и предварительным идентификационным исследованием по голосу и речи.

Тимофеев А.В.

*Доктор технических наук, ООО «Центр речевых технологий»,
г. С.-Петербург.*