

МЕТОДЫ ВЫДЕЛЕНИЯ ИНФОРМАЦИОННЫХ ПРИЗНАКОВ РЕЧЕВЫХ СИГНАЛОВ

Езовит А. В., Гуринович А. Б.

Кафедра информационных технологий автоматизированных систем,
Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: ezovit99@yandex.ru, gurinovich@bsuir.by

В настоящее время важной задачей в области речевой обработки является выделение информационных признаков сигнала. Достижения в данной сфере могут применяться для идентификации личности, её описания, сбора информации о физическом состоянии и роде деятельности. В данной статье приведены методы, которые могут использоваться для решения задачи выделения информационных признаков.

ВВЕДЕНИЕ

Важной задачей в области речевой обработки является выделение информационных признаков сигнала. Информационными признаками являются характеристики голоса и речи, которые уникальны для каждого человека. Их выделение из сигнала позволит составить «портрет» речи человека, позволяющий определить кем он является и при необходимости идентифицировать его. Примерами информационных признаков в голосовом сигнале (обработка голоса) являются основной тон (характеризующий гендер человека) и обертона (характеризующий спикера)[1]. В речевом сигнале — темп и благозвучность речи, которые, в отличие от приведенных выше голосовых признаков, не обоснованы (или обоснованы не полностью) анатомией каждого конкретного человека и являются следствием привычек и тренировок[2].

Задачей, схожей с выделением признаков, является задача идентификации человека. При решении этой задачи собирается вектор признаков, являющийся «слепок голоса» и представляющий из себя смесь из разных признаков. Таким образом, цель задачи по выделению — разделение этой смеси на составляющие ее информационные признаки.

Задачи по выделению признаков актуальны, например, для подтверждения личности клиентов банков, установления личности в криминалистике и других областях.

I. ХАРАКТЕРИСТИКИ ГОЛОСА

Первоначально для выполнения задачи выделения информационных признаков в речевом сигнале определяют, что может служить определяющей характеристикой голоса. Признаки, заключенные в речевом сигнале, делятся на две группы: низкоуровневые/голосовые (признаки, характеризующие голос и обусловленные анатомией человека) и высокоуровневые/речевые (признаки, характеризующие речь, связанные с манерой общения)[3].

К низкоуровневым относятся:

- Сила голоса (амплитуда)

- Высота голоса (частота)
- Тембр голоса (окраска голоса, задаваемая обертонами)
- Полетность (слышимость голоса на большом удалении от источника)
- Благозвучность голоса (звонкость и частота звучания без призвуков)

К высокоуровневым относятся:

- Подвижность (способность человека к изменению параметров голоса)
- Тон голоса (эмоциональная окраска)
- Темп речи (скорость, с которой говорит человек)
- Артикуляция и дикция (отвечают за четкость произнесения звуков в соответствии с нормами языка)

II. ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА ИСХОДНОГО СИГНАЛА

Для выделения информационных признаков в речевом сигнале проводится анализ отцифрованного сигнала. Анализ может проводиться непосредственно над амплитудно-временным представлением сигнала или над его спектральным видом. Процесс обработки и приведения к спектральной форме производится в несколько этапов:

1. Фильтрация сигнала от выбросов, вызванных микрофоном (данные выбросы зависят от качества микрофона и могут поднимать средний уровень сигнала на некоторое значение, что в свою очередь приводит к проблемам обработки) и низкочастотных шумов (данные шумы особо выделяются в местах спада энергии в амплитудно-временной области из-за чего могут быть восприняты за голосовую активность и вызвать проблемы при анализе).
2. Нормализация сигнала (процесс выравнивания громкости сигнала относительно какого-либо эталона, может применяться пиковая и средняя нормализация, определяющая будет с эталоном сравниваться пи-

ковое или среднее значение соответствующе).

3. Разбиение сигнала на временные кадры (длина кадра выбирается в зависимости от длины интервала времени, на котором происходит всплеск мощности, и обычно находится в диапазоне 10-150 мс).
4. Применение весовой функции (например окно Хэмминга) для уменьшения искажений (при этом выделяется центральная часть кадра и происходит плавное затухание по краям, что упрощает дальнейшую обработку).
5. Осуществления преобразования Фурье.

После окончания преобразования осуществляется переход в логарифмический масштаб, к которому чувствительно человеческое ухо (пример Мэл-шкала, данная шкала имеет линейную природу на диапазоне до 1000 Гц и логарифмическую для частот выше).

Первичная обработка сигнала является важным шагом на пути к последующему анализу, поскольку правильно проведенная обработка может существенно снизить мощности необходимые на анализ.

III. МЕТОДЫ ВЫДЕЛЕНИЯ ПРИЗНАКОВ

При решении задачи распознавания спикера, по полученному спектру собирается суммарный вектор признаков, для которого в последующем на основе метрики (определяется расстояние между векторами, например евклидово расстояние, L1-норма, расстояние Махалобиса) или оценки вероятности (модель гауссовских смесей) определяется схожесть с имеющимся вектором (характеризующим голос спикера) и делается вывод, являются ли говорящие одним человеком[4].

Для эффективного решения задачи выделения признаков предлагается следующее: аналогичным образом собирается вектор признаков, однако в последующем данный набор не сравнивается с уже имеющимся вектором, а аналогичными методом определения метрического расстояния или оценки вероятности проверяется входимость признака в ранее определенную группу. Например, при предварительной подго-

товке выборка из мужских, женских и детских голосов будет обработана и проанализирована, после чего для каждого голоса будет выделен свой вектор признаков, на основе которых произведется сравнение, определяющее группы (интервал), входимость в которую означает, наличие определенного признака в образце голоса (в данном примере мужской, женский, детский). Подобным подходом может оцениваться эмоциональность в голосе, сила, чистота звучания и другие признаки.

Эффективно будет добавить еще один выделяемый признак – скорость произношения, которая может отличаться от человека к человеку. Оценку данного признака можно произвести на основе спада амплитуд во временной или частотной области.

IV. ЗАКЛЮЧЕНИЕ

Предложенные методы позволят провести анализ голоса спикера на входимость его в различные группы, что в свою очередь позволяет выделять определенные информационные признаки. Очевидно, что при составлении групп, сигнализирующих о наличии того или иного признака, важным моментом, влияющим на правильность определения, является выборка заранее обработанных голосов, на основе которых строится интервал группы. Результаты выделения в совокупности с моделями, построенными в других областях знаний, позволят не только описать голос человека, но и установить другие закономерности. Например, физическое и эмоциональное состояние или возможный род деятельности. Это позволит эффективно решать проблему выделения информационных признаков речевого сигнала.

V. СПИСОК ЛИТЕРАТУРЫ

1. Григорьев Е.И., Тычинина В.М. Звуки речи и их коммуникативная функция : учебное пособие для студентов филологических специальностей, аспирантов и преподавателей / Е.И. Григорьев, В.М. Тычинина. — Тамбов, ТГУ им. Г.Р.Державина. — 84 с.
2. 4brain [Электронный ресурс]. – Москва, 2021 – Режим доступа: <https://4brain.ru/voice/properties.php> – Дата доступа: 02.11.2021.
3. Е.А. Первушин обзор основных методов распознавания дикторов // Математические структуры и моделирование 2011, вып 24: Омск 2011. – С. 41-54.
4. habr [Электронный ресурс]. – Москва, 2021 – Режим доступа: <https://habr.com/ru/post/336516/> – Дата доступа: 02.11.2021.