

РАЗРАБОТКА СИСТЕМЫ РАСПОЗНАВАНИЯ ДИКТОРА НА ОСНОВЕ МОДЕЛИ ДУРЛАХА

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Кручок Д.Н.

Петровский А.А. – д.т.н., профессор

Системы распознавания речи и диктора на сегодняшний день активно развиваются. Результаты работы в этих областях уже внедряются и используются в различных областях (например, системы обеспечения доступа, банковская сфера). Показатели распознавания систем, которые тестируются в условиях агрессивной акустической обстановки, быстро деградируют по мере уменьшения соотношения сигнал-шум (SNR), тем самым ставят задачу повысить эффективность систем распознавания диктора и речи в условиях акустических шумов.

Бинауральная демаскировка – эффект усиления полезного сигнала слуховой системой человека, при условии, что в одно ухо поступает шум, а в другое сигнал и шум (при этом шум полностью маскирует полезный сигнал). Это срабатывает, когда в сигнале присутствуют низкочастотные составляющие. Модель работы слуховой системы, учитывающая этот эффект, – модель Дурлаха, или «уравнивание-сокращение». [1].

Принцип работы модели «уравнивание-сокращение» состоит в том, что звуковой сигнал, пройдя через уши, попадёт в мозг как моноуральным, так и бинауральным путями. Т.е. мозг, выбирает сигнал с подходящим уровнем SNR между тремя каналами, два моноуральных и один бинауральный. Моноуральные каналы идут напрямую к мозгу, а бинауральный ещё включает в себя стадию нормировки – уравнивания амплитуд сигналов с обеих ушей, и стадию сокращения – когда эти сигналы вычитаются друг из друга.

К плюсам данной модели можно отнести то, что полезный сигнал усиливается на фоне шума, что делает систему распознавания более устойчивой к акустическим шумам. К минусам относится необходимость использования двух микрофонов, либо клонировать сигнал и дополнительно его обрабатывать для достижения эффекта бинауральной записи.

Данную модель можно использовать при построении системы распознавания диктора [2]. Предлагается использовать не один микрофон, а три микрофона. Два основных будут использоваться для получения сообщений от дикторов и расположены недалеко друг от друга (примерно 18 см – среднее расстояние между ушными каналами) и направлены в разные стороны, а третий микрофон – для снятия окружающих акустических шумов – будет находиться подальше.

Процесс обработки речевого сигнала на основе модели Дурлаха, изображен на рисунке 1.

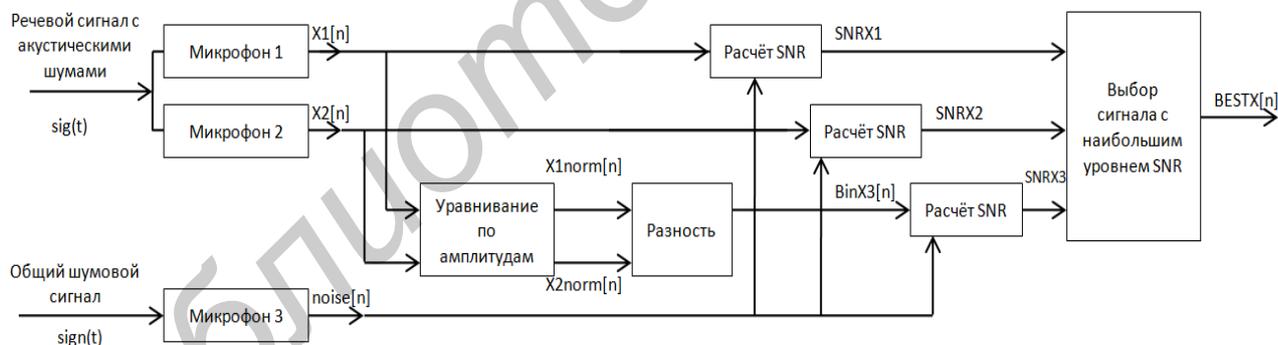


Рис. 1 – Схема обработки сигнала

В результате обработки будет получен сигнал с наибольшим уровнем – SNR BESTX[n], который в дальнейшем будет использоваться в системе распознавания.

Были проведены эксперименты по предложенной обработке сигнала. Использовались два микрофона (третьего не было в наличии) на небольшом расстоянии друг от друга (15-20 см), на них была записана речевая фраза в помещении с небольшими шумами. Потом к сигналам был добавлен аддитивно белый шум. Полученные уровни SNR для трёх сигналов: SNRX1 = 1,3853 дБ, SNRX2 = 0,4962 дБ, SNRX3 = 2,5128 дБ. Данный результат подчёркивает, что использование эффекта бинауральной демаскировки целесообразно для задач распознавания диктора.

Таким образом, использование двух микрофонов в процессе записи сигнала позволяет получить третий сигнал, используя обработку по модели Дурлаха, с наибольшим уровнем отношения сигнал-шум, а значит, улучшит результат дальнейшей идентификации диктора.

Список использованных источников:

1. Алдошина И.А. Основы психоакустики / И.А. Алдошина – М.: Оборонгиз, 2000. – С.34-35.
2. Кручок Д.Н., А.А. Петровский Система идентификации диктора в акустических шумах на основе антропоморфической обработки речевого сигнала / Д.Н. Кручок, А.А. Петровский // Доклады БГУИР, – 2016. – С.47-53.