

ПСИХОАКУСТИЧЕСКИ МОТИВИРОВАННЫЙ АЛГОРИТМ ЭКСТРАКЦИИ ХАРАКТЕРИСТИЧЕСКОГО ВЕКТОРА ДЛЯ ИДЕНТИФИКАЦИИ ДИКТОРА В АКУСТИЧЕСКИХ ШУМАХ

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

Кручок Д. Н.

Петровский А. А. – д-р. техн. наук, профессор

Речевые технологии находят все большее распространение во сферах жизни человека: охранные системы и системы безопасности, банковские операции, концепция «умный дом», робототехника, системы управления оборудованием, средствах телекоммуникаций. Хотя и эффективность работы систем распознавания в закрытых или помещениях с благоприятной акустической обстановкой находится на довольно высоком уровне, результаты распознавания данных систем на улице или в местах с наличием акустических шумов ещё далеки от требуемых. В работе предлагается психоакустически мотивированный алгоритм экстракции характеристического вектора для его дальнейшего использования в идентификации диктора в акустических шумах.

Самой лучшей системой распознавания на сегодняшний день остаётся человек [1]. Использование особенностей строения слуховой системы человека в автоматических системах распознавания (АСР) повышает эффективность работы данных систем, позволяет эксплуатировать их в акустически неблагоприятных условиях (с наличием шума).

Психоакустически мотивированный алгоритм экстракции – алгоритм, который использует особенности строения человеческой слуховой системы для извлечения характеристических признаков из звукового сигнала. В основе данного алгоритма лежит слуховое преобразование. Оно моделирует распространение бегущей волны в улитке уха человека, где звуковой сигнал раскладывается в набор подсигналов с разной частотой. По этой причине данный банк фильтров называют кохлеарным (cochlea – улитка). Рассматриваемый алгоритм состоит из этапов, представленных на рисунке 1, и частично моделирует процесс обработки звукового сигнала слуховой системой человека.



Рис.1 – Сравнение этапов обработки сигнала: а) схема рассматриваемого алгоритма экстракции; б) процесс обработки звукового сигнала в слуховой системе

После слухового преобразования получается сигнал, разложенный на частотные полосы, что моделирует работу базилярной мембраны (БМ) в улитке уха человека. Таким образом, если $f(t)$ – исходный сигнал, а импульсная характеристика БМ – $\psi_{a,b}(t)$ [2], то его слуховое преобразование будет иметь следу-

$$\text{ющий вид: } T(a,b) = \int_{-\infty}^{\infty} f(t)\psi_{a,b}(t)dt$$

Для моделирования поведения волосковых клеток, используется следующая функция:

$$h(a,b) = T(a,b)^2; \forall T(a,b),$$

где $T(a,b)$ – выходные значения банка фильтров. Предполагается, что все другие детальные функции в процессе обработки сигнала слуховой системой человека были проигнорированы или включены в

акустические характеристики фильтра. На следующем этапе, значения функции $h(a, b)$ волосковых клеток для каждой полосы усредняются, при этом используется окно разной длины в отличие от дискретного преобразования Фурье, где размер окна фиксирован. Чем выше частота, тем короче окно – это предотвращает от сглаживания информации длинным окном.

$$S(i, j) = 1/d \sum_{b=l}^{l+d-1} h(i, b), l = 1, L, 2L, \dots; \forall i, j$$

где $d = \max\{3, 5\tau_i; 20\text{мс}\}$ – длина окна; τ_i – период центральной полосы i -й полосы и $L = 10\text{мс}$ – размер окна сдвига.

Для учёта нелинейности были использованы веса функции громкости [2]:

$$y(i, j) = S(i, j)^{1/3}.$$

И на последнем этапе применяется дискретное косинусное преобразование (ДКП), для уменьшения размера полученных признаков. Эти признаки получили название кохлеарные кепстральные коэффициенты (ККК) и они являются тем пространством, в котором будет происходить идентификация диктора.

Для экспериментальной оценки признаков ККК для текстонезависимой идентификации в акустических шумах, был спроектирован идентификатор дикторов на основе нейронных сетей. Также было произведено сравнение результатов распознавания ККК признаков с существующими мел-частотными кепстральными коэффициентами (МКК) и характеристиками, полученные с помощью алгоритма RASTA (Relative Spectral Transform – Относительное спектральное преобразование) в задаче идентификации диктора в акустических шумах. В качестве базы была составлена речевая база, основанная на базе SSC[3]. Было подготовлено 3 множества: обучающее, настраиваемое и тестовое. Сравнение результатов идентификации диктора с различными уровнями шумов (представленный тип шума – bubble шум) представлено на рисунке 2.

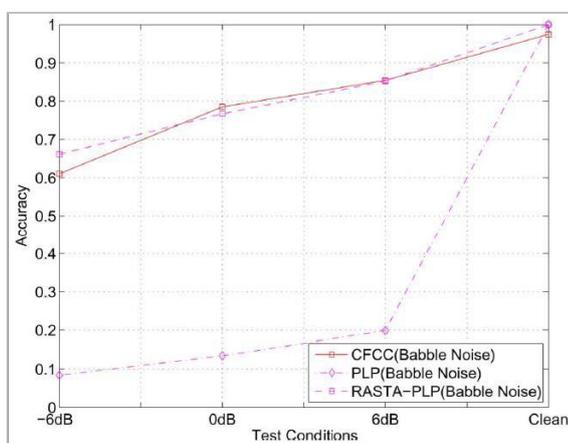


Рис.2 – Сравнение результатов распознавания

Проведённые эксперименты демонстрируют, что в условиях наличия акустических шумов различного типа (белый шум, bubble шум, машинный шум), рассмотренный алгоритм экстракции характеристического вектора справляется с задачей идентификации значительно лучше, чем алгоритмы извлечения МКК и относительного спектрального преобразования при наличии шума любого рассматриваемого типа.

Список использованных источников:

1. The PASCAL CHiME speech separation and recognition challenge [Электронный ресурс] – Электронные данные. – Режим доступа : <http://spandh.dcs.shef.ac.uk/projects/chime/PCC/results.html>.
2. Q. Li, An auditory-based feature extraction algorithm for robust speaker identification under mismatched conditions / Proc. IEEE Transactions Audio, Speech, and Language Processing., – New Paltz, NY, 2011.
3. The PASCAL CHiME speech separation and recognition challenge [Электронный ресурс] – Электронные данные. – Режим доступа : <http://spandh.dcs.shef.ac.uk/projects/chime/PCC/datasets.html>.

КОМБИНИРОВАННАЯ СИСТЕМА УЛУЧШЕНИЯ РАЗБОРЧИВОСТИ АУДИО СИГНАЛА В АГРЕССИВНОЙ ШУМОВОЙ СРЕДЕ

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

Белый В.Л.

Петровский А.А. – д-р. техн. наук, профессор

Современные устройства и технологии позволяют передавать, принимать и воспроизводить аудио сигналы, находясь в различных условиях окружающей среды. Это приводит к возникновению проблемы разборчивости принятого,