

Как видно из экспериментальных результатов, приведенный способ формирования высокочастотной составляющей спектра сигнала даёт прирост качества восстановленного сигнала. При этом, в качестве дополнительных данных, которые необходимо передать декодеру для реконструкции выступает энергия в целевой частотной полосе, т.е. один параметр на полосу. В проведенных экспериментах перенос осуществлялся в три полосы, следовательно на каждый фрейм требовалось только три дополнительных параметра, что незначительно увеличит битрейт, при этом позволит получить более высокое качество выходного сигнала.

Список использованных источников:

1. Petrovsky, Al. Scalable parametric audio coder using sparse approximation with frame-to-frame perceptually optimized wavelet packet based dictionary / Al. Petrovsky, V. Herasimovich, A. Petrovsky // AES 138th Convention, paper 9264 – Warsaw, Poland, 2015, May 7 - 10.
2. Zwicker, E. Psychoacoustics: facts and models / E. Zwicker, H. Fastl. – Berlin: Springer-Verlag, 1990. – 354 p.
3. Valin, J.-M. High-Quality, Low-Delay Music Coding in the Opus Codec / J.-M. Valin, G. Maxwell, T. Terriberry, K. Vos // AES 135th Convention, paper 8942 – New York, USA, 2013, October 17 - 20.
4. Dietz, M. Spectral Band Replication, a novel approach in audio coding / M. Dietz, L. Liljeryd, K. Kjörling, O. Kunz // AES 112th Convention, paper 5553 – Munich, Germany, 2002, May 10 - 13.
5. Huber, R. PEMO-Q – A New Method for Objective Audio Quality Assessment Using a Model of Auditory Perception / R. Huber, B. Kollmeier // IEEE Transactions on audio, speech, and language processing, vol. 14. – November, 2006. – pp. 1902-1911.

СИСТЕМА ИДЕНТИФИКАЦИИ ДИКТОРА В АКУСТИЧЕСКИХ ШУМАХ НА ОСНОВЕ АНТРОМОРФИЧЕСКОЙ ОБРАБОТКИ РЕЧЕВОГО СИГНАЛА

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

Кручок Д. Н.

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

Рассматривается система идентификации диктора в акустических шумах с использованием антропоморфической обработки речевого сигнала. Полученный характеристический вектор на базе данного преобразования используется в качестве признаков для системы идентификации диктора. В качестве решающих правил применяются нейронные сети прямого распространения. Приводятся результаты распознавания разработанной системы идентификации диктора.

В работе [1] был предложен алгоритм получения характеристического вектора, который пытается смоделировать процессы обработки речевого сигнала человеческим ухом. Рассматриваемый алгоритм состоит из этапов, представленных на рис. 1, а 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 человека

Полученные кохлеарные кепстральные коэффициенты (ККК) используются в системе идентификации в качестве признаков. В качестве решающих правил системы идентификации диктора наиболее часто используются следующие методы: векторное квантование, гауссовские смеси, нейронные сети и метод опорных векторов. В данном исследовании были выбраны нейронные сети прямого распространения для получения сравнительных результатов распознавания диктора с результатами, представленными в работе [1], в которой были использованы модели гауссовых смесей. Для экспериментальной оценки системы идентификации диктора в акустических шумах, был спроектирован идентификатор дикторов на основе нейронных сетей прямого распространения. Модель системы распознавания представлена на рис. 2 и имеет два режима работы: обучение и идентификация. Архитектура используемой сети выбиралась постепенно от

простого однослойного персептрона к многослойным экспериментально. Используемые типы нейронных сетей: многослойный персептрон (два и три слоя): 2 слоя: 56 входных – 19 выходных. 3 слоя: 56 входных – 30 промежуточных – 19 выходных. Функция активации на всех слоях, кроме выходного: сигмоид, на выходном – функция *softmax*. Процесс обучения происходит с учителем.

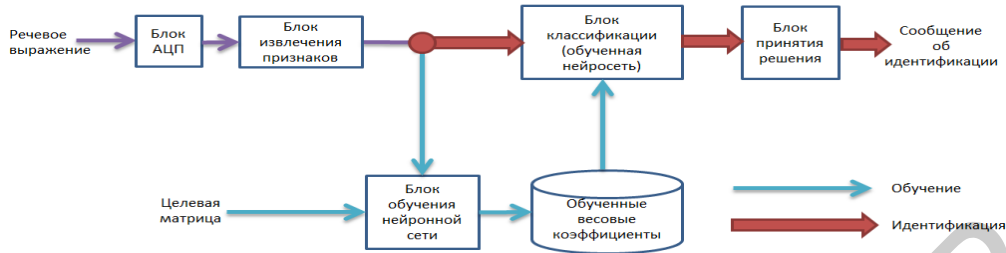


Рис. 2. – Модель системы распознавания

Используемая речевая база была подготовлена на основе базы, сформированной для соревнований по разделению и распознаванию речи и дикторов в условиях с шумами [2]. Было выбрано 4 диктора, у каждого по 18 речевых фраз средней продолжительностью около 2 секунд, не содержащих шумы.

В алгоритме экстракции характеристического вектора использовалось 32 кохлеарных фильтра [1]. Оценкой системы идентификации выступает точность распознавания диктора – отношение числа правильно распознанных речевых выражений к общему числу высказываний, участвовавших в распознавании. Данная оценка выражается в долях единицы или в процентах. Для получения сравнительных результатов был реализован алгоритм получения кепстральных коэффициентов в шкале Мел (мел-частотные кепстральные коэффициенты – МЧКК) и использован в системе идентификации диктора. Система идентификации реализована на языке Matlab. Результаты распознавания на обучающем множестве представлены на рис. 3. Тип используемого шума – лепет, бормотание (англ. «*babble*»).

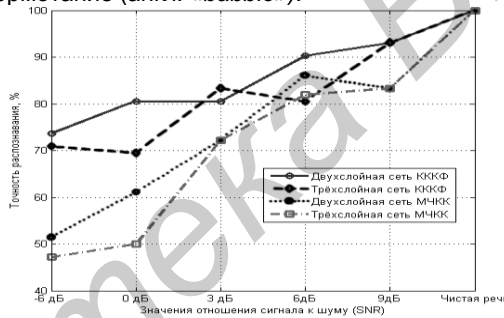


Рис. 3. Сравнение результатов для обучающего множества (тип шума – бормотание)

Как видно из рисунка 4 в условиях без шумов признаки ККК показывают сравнимые результаты с признаками МЧКК и имеют 100% точность распознавания (на ограниченном числе выражений). По мере того, как увеличивается уровень шума, точность системы распознавания падает, однако точность идентификации с использованием признаков ККК значительно лучше, чем МЧКК. Например, при SNR равным 0 дБ, точность распознавания для признаков МЧКК составляет 62 %, а для ККК – 80 %. На тестовом множестве точность распознавания всей системы снижается, однако результаты идентификации с использованием признаков ККК превосходят результаты МЧКК: 77% к 69% при SNR равным 0 дБ (рис. 4). Эффективность используемого алгоритма экстракции характеристического вектора в системе идентификации на нейронных сетях при уровне SNR -6 дБ больше на 20% для МЧКК, и на 5-10% для ККК, чем в системе на основе гауссовых смесей [1].

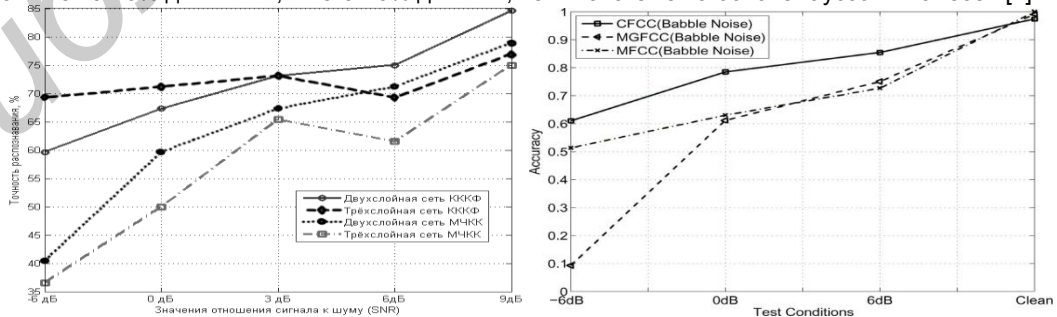


Рис. 4. Результаты распознавания для тестового множества (слева – система идентификации на нейронных сетях, справа – система идентификации на гауссовых смесях [1]) для типа шума – бормотание

Список использованных источников:

1. Q. Li, An auditory-based feature extraction algorithm for robust speaker identification under mismatched conditions / IEEE Transactions Audio, Speech, and Language Processing., – 2011. – Vol. 19, № 6. – P.1791-1801.
2. The PASCAL CHiME speech separation and recognition challenge [Electronic resource]. – 2013. – Mode of access :

СИСТЕМА УЛУЧШЕНИЯ ВОСПРИЯТИЯ И 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

В современных системах улучшения восприятия и разборчивости аудио сигнала в агрессивной шумовой среде и в системах шумоподавления отдельную нишу занимают приложения, предназначенные для мобильных и портативных устройств, предполагающих использования наушников.

Универсальная система улучшения восприятия и разборчивости аудио сигнала требует достижения высокого уровня шумоподавления для шумов различной природы, различного уровня и различного частотного состава. Исходя из данных требований было принято решения разработать систему шумоподавления, представляющую собой двухканальную интеграцию субполосного аудио интегрированного алгоритма активного шумоподавления (ActiveNoiseControl – ANC) с маскированием, основанным на принципах психоакустического анализа.

На рисунках 1 и 2 приведены выбранные структуры субполосного аудио интегрированного алгоритма активного шумоподавления (ANC) и модели расчета коэффициентов усиления маскиера, базирующейся на принципах психоакустики:

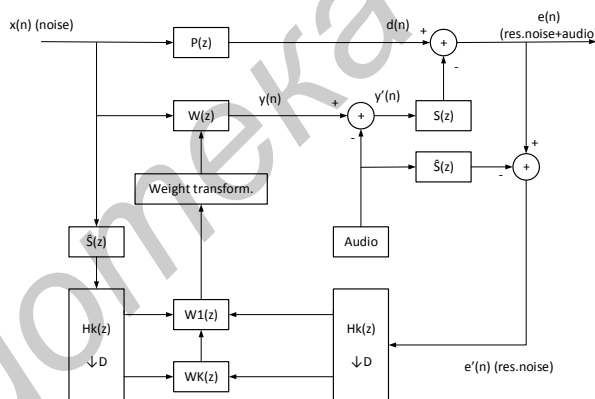


Рис. 1. Структурная схема ANC

Работу субполосного аудио интегрированного алгоритма активного шумоподавления (ANC) можно представить следубщими формулами:

$$y_n = \sum_{n=0}^{N-1} w_n x(n-N)$$

$$y'_n = y_n - a(n)$$

$$e_n = d_n - y'_n s'(n)$$

$$e'_n = e_n - a'(n)$$

Где, $a(n)$ – целевой аудио сигнал; $y'_n s'(n)$ – $y'(n)$ после прохождения вторичного пути; $a'(n)$ – $a(n)$ после прожеля вторичного пути, N – порядок широкополосного адаптивного фильтра;

В качестве алгоритма адаптации коэффициентов субполосных фильтров использовался алгоритм FxLMS:

$$w_k(n+1) = w_k(n) + \mu \frac{X_n e(n)}{\epsilon + X(n)^2}$$