

На рисунках 3 приведена структурная схема алгоритма маскирования:

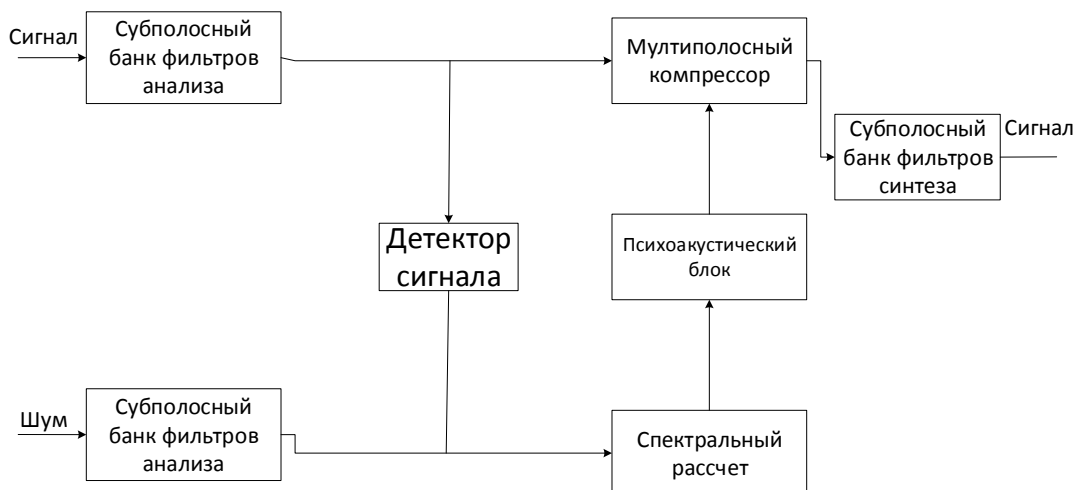


Рис. 3 – Структурная схема алгоритма маскирования

Для создания математических моделей был использован язык технических вычислений MatLab. Для проектирования прототипа устройства была использована процессорная система на базе ARM-ядра.

В результате построена система улучшения разборчивости аудио сигнала в агрессивной шумовой среде. Основными преимуществами является двуэтапность обработки, которая позволяет применять данную систему в различных шумовых средах, для различных аудио сигналов и препятствует достижению критически опасного уровня аудио сигнала для слуховой системы человека, также данная система может работать для передачи аудио сигнала, подавляя шумы в сигнале.

Список использованных источников:

1. Конг-Аик Ли. Субполосная адаптивная фильтрация, теория и применение / Конг-Аик Ли, Вун-Сенг Ган, Сен М. Куо. // – JohnWileyandSons, Ltd, 2009. – 324 с.
2. Петровский, А. А. Цифровые банки фильтров: анализ, синтез и применение в мультимедиа системах / А. А. Карпушкин, М. Парфенюк, А. Борович, М.З. Лившиц // Уч. метод. пособие для студентов учреждений, обеспечивающих получение высшего образования по специальности "Электронные вычислительные средства". – Минск, 2006. – 81 с.
3. Bjarnason, E. Analysis of filtered-X LMS algorithm. / Bjarnason E. // - IEEE Trans. Speech audio Process., 3(6), 1995, 504-514.
4. Davari, P. Designing a new robust on-line secondary path modeling technique for feedforward active noise control systems/ Davari P, Hassanpour H. // - Signal processing, 89(6), 2009, 1195 – 1204.
5. Rees, L. Psychoacoustic modelling for active noise control system/ Rees L., Elliot S.J.// - Proc. Inst. Acoust., 26(pt.2), 2004.

СЛУХОВОЕ ПРЕОБРАЗОВАНИЕ. ЕГО ИСПОЛЬЗОВАНИЕ В ИДЕНТИФИКАЦИИ ДИКТОРА

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

Кручок Д. Н.

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

В современных автоматических системах распознавания (АСР) одной из главных задач является обеспечение распознавания в условиях с различной акустической обстановкой. Для достижения высоких показателей эффективности распознавания в акустических шумах, целесообразно использовать алгоритмы, моделирующие процесс обработки звукового сигнала слуховой системой человека. Основой для таких алгоритмов служат различные преобразования, одним из которых является слуховое преобразование.

Слуховое преобразование – преобразование, которое переводит сигнал из временной области в набор выходов банка фильтров. Частотные характеристики и распределение полос банка фильтров такие же, как и у базилярной мембраны в улитке в среднем ухе человека. Когда звук попадает в человеческое ухо, акустическая энергия с внешнего уха превращается в механическую энергию через среднее ухо, представленное тремя косточками. Когда последняя косточка, стремечко, колеблется, то это приводит в движение жидкость внутри улитки, создавая бегущие волны на базилярной мембране (БМ).

Импульсную характеристику БМ в улитке можно представить, как:

$$\psi_{a,b}(t) = 1/\sqrt{|a|} \cdot \left(\frac{t-b}{a}\right)^\alpha \cdot \exp\left[-2\pi f_L \beta \left(\frac{t-b}{a}\right)\right] \cdot \cos\left[2\pi f_L \beta \left(\frac{t-b}{a}\right) + \theta\right] \cdot u(t-b),$$

где $1/\sqrt{|a|}$ – коэффициент нормализации энергии; a, b – параметры, определяющие расширение шкалы и временной сдвиг соответственно; $\alpha > 0, \beta > 0$ – определяют форму и ширину кохлеарного фильтра в частотной области; $u(t)$ – ступенчатая функция (1 при $t \geq 0, 0$ – в ост. случаях); θ – коэффициент, должен выбираться с учетом условия $\int_{-\infty}^{\infty} \psi(t) dt = 0$. Параметр a можно определить, зная центральную частоту текущего фильтра f_c и начальную частоту f_L во всем банке фильтров: $a = f_c / f_L$.

Таким образом, если $f(t)$ – исходный сигнал, то его слуховое преобразование будет иметь следующий вид:

$$T(a,b) = \int_{-\infty}^{\infty} f(t) \psi_{a,b}(t) dt.$$

Изменяя параметры α, β можно получить различные форму и ширину для каждого фильтра в кохлеарном банке фильтров, которые будут наиболее точно соответствовать модели слуховой системы человека. На рисунке 1 представлены частотные характеристики 5-ти первых фильтров при различных значениях коэффициента β и $\alpha = 0,8$. Указанные значения были выбраны как наиболее оптимальные для анализа речевого сигнала для дальнейшей идентификации диктора.

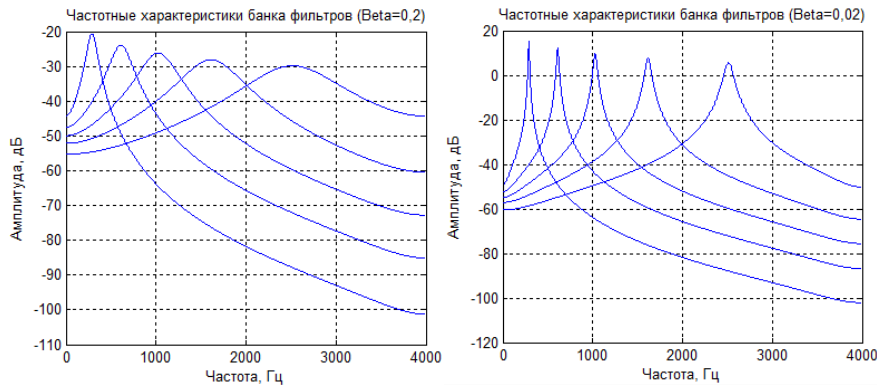


Рис. 1 – Частотные характеристики банка фильтров

Импульсные характеристики базилярной мембраны на разных частотах показаны на рисунке 2, при этом $\alpha = 0,8; \beta = 0,02$. Метки по оси ординат слева от каждого графика соответствуют центральным частотам банка фильтров в Гц.

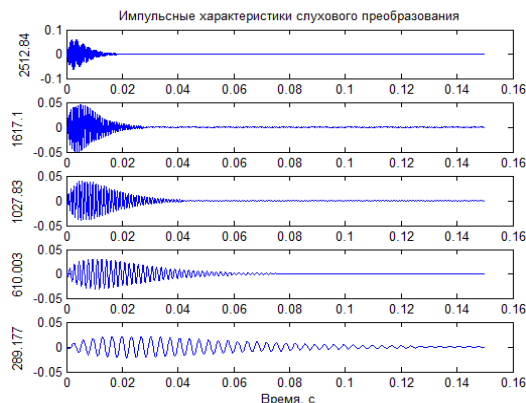


Рис.2 – Импульсные характеристики БМ для слухового преобразования

Такие результаты очень похожи на результаты физиологических экспериментов [1], а это в свою очередь, свидетельствует о том, что данное преобразование позволяет получить характеристический вектор признаков, который будет соответствовать характеристикам, которые выделяет слуховая система человека.

Спектрограмма, полученная с помощью слухового преобразования, показана на рисунке 3. В качестве тестового сигнала был взят речевой сигнал с частотой дискретизации 8 кГц; параметры имеют следу-

ющие значения: $\alpha = 0,8$; $\beta = 0,2$; количество используемых фильтров в банке равно 32. Фраза на русском языке: «штурман просил продолжать разворот».

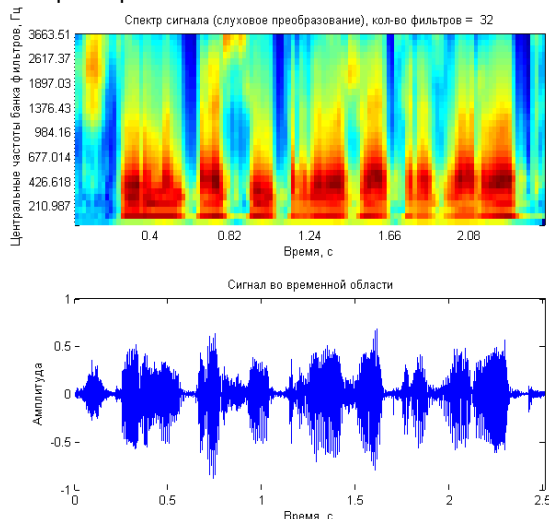


Рис.3 – Спектр сигнала, полученный с помощью слухового преобразования

Данное преобразование моделирует импульсные характеристики базилярной мембраны и её особенности нелинейного распределения частоты. Рассмотренное преобразование имеет следующие преимущества: устойчивость к шумовой составляющей сигнала, не имеет четко выраженных гармоник и также не даёт вычислительного шума. Эти преимущества могут быть использованы в различных приложениях, как в задачах идентификации диктора для получения характеристического вектора, так и для шумоподавления, синтеза музыки и речи и т.д. В области идентификации диктора, результатом слухового преобразования является сигнал, разложенный на частотные полосы, так как это делает БМ. Далее по аналогии со слуховой системой человека моделируется поведение волосковых клеток (внутренних и внешних) и учёт их нелинейности восприятия. После, следует этап уменьшения размерности полученных характеристик. Полученный характеристический вектор, называют кохлеарными кепстральными коэффициентами (ККК).

Таким образом, было рассмотрено и реализовано с помощью среды MATLAB слуховое преобразование. Были подобраны наиболее оптимальные параметры используемого банка кохлеарных фильтров для их дальнейшего использования в задачах распознавания. Полученные с помощью слухового преобразования характеристические признаки будут более устойчивы к изменению акустической среды в процессе распознавания, что повысит эффективность работы систем распознавания в акустических шумах.

Список использованных источников:

1. B. S. Atal, "Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification," J. Acoust. Soc. Amer., vol. 55, pp. 1304–1312, 1974.
2. Q. Li, An auditory-based transform for audio signal processing / Proc. IEEE Workshop Applicat. Signal Process. Audio Acoust., – New Paltz, NY, 2009.

ОБУЧЕНИЕ РЕКУРРЕНТНОЙ НЕЙРОННОЙ СЕТИ ДЛЯ РАСПОЗНАВАНИЯ РЕЧЕВЫХ КОМАНД

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

Жук И.Н.

Азаров И.С. – к-т. техн. наук, доцент

В современных автоматизированных системах взаимодействия между человеком и вычислительной техникой все чаще встречается голосовая составляющая. Голосовое управление во многом позволяет убрать преграды непосредственного механического взаимодействия с устройствами. Реализация голосового командного интерфейса производится в большинстве случаев с помощью различных нейронных сетей.

Применение рекуррентной нейронной сети подразумевает использование выходных значений сети в качестве входных, что позволяет сохранять внутри сети информацию о предыдущих входных значениях. Данный подход позволяет представлять звук естественным последовательным образом. Подобная структура сети предполагает использование соответствующих обучающих алгоритмов. Для изучения был выбран алгоритм обратного распространения ошибки во времени. Суть алгоритма заключается в том, что мы разворачиваем рекуррентную сеть в очень длинную обычную сеть прямого распространения.