

УДК 534.773

СИСТЕМА ПОВЫШЕНИЯ РАЗБОРЧИВОСТИ РЕЧИ НА ОСНОВЕ МОДЕЛИ СЛУХОВОГО ВОСПРИЯТИЯ И КОМПРЕССИИ ДИНАМИЧЕСКОГО ДИАПАЗОНА СИГНАЛА

Семенюк Н.А., студент гр.750701

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

Порхун М.И. – магистр техн. наук

Аннотация. В работе рассмотрена система повышения разборчивости речи на основе модели слухового восприятия и компрессии динамического диапазона сигнала. Функционирование системы сводится к субполосной обработке звукового сигнала, согласованной с механизмом работы слуховой системы человека. Банк гамматон-фильтров используется для разложения звукового сигнала на компоненты (субполосы). Каждая субполоса проходит через свой компрессор динамического диапазона, задача которого – автоматический выбор коэффициента усиления для текущей субполосы. Усиленные субполосы синтезируются в выходной сигнал путём суммирования. Настройка системы выполняется по аудиограмме тугоухого человека. Для проверки работоспособности системы проведено MATLAB-моделирование.

Ключевые слова. Повышение разборчивости речи, гамматон-фильтр, компрессор динамического диапазона, аудиограмма.

По состоянию на 2021 год более 5% населения мира (430 миллионов человек), нуждаются в реабилитации для решения проблемы потери слуха. По прогнозам, к 2050 г. по меньшей мере 700 миллионов человек будут нуждаться в реабилитационных услугах в связи с потерей слуха и около 2,5 миллиарда человек в той или иной степени будут страдать от проблем со слухом [1]. В связи с этим весьма актуальна задача оперативной коррекции слуховых патологий.

В данной работе рассмотрена система повышения разборчивости речи на основе модели слухового восприятия и компрессии динамического диапазона сигнала. Структура системы представлена на рисунке 1.

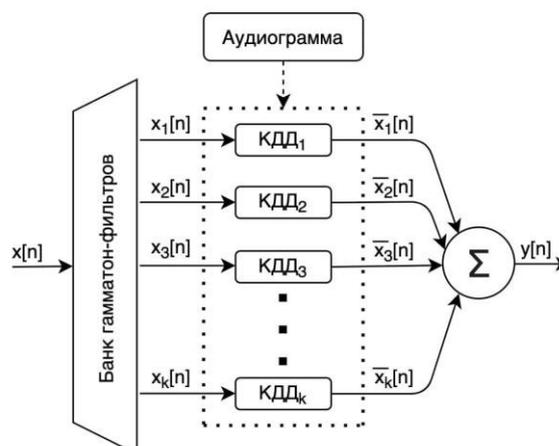


Рисунок 1 – Структура системы повышения разборчивости речи

Входной сигнал $x[n]$ разделяется на субполосы при помощи 20-канального банка гамматон-фильтров [2]. Каждый субполосный сигнал проходит через отдельный компрессор динамического диапазона (КДД). Основная задача компрессоров – расчёт коэффициентов субполосного усиления сигнала. Система настраивается по аудиограмме слабослышащего человека. Выходной звуковой сигнал $y[n]$ синтезируется при помощи суммирования обработанных субполос.

В качестве модуля, разделяющего сигнал на субполосы, используется банк гамматон-фильтров, имитирующий механизм частотной декомпозиции звука в улитке [2]. Импульсная характеристика банка гамматон-фильтров описывается следующим выражением:

$$h_k(t) = e^{-\alpha t} e^{-2\pi i f_c t} \cos(2\pi f_c t + \phi), \text{ при } t \geq 0 \quad (1)$$

где α – порядок фильтра, f_c – эквивалентная прямоугольная полоса пропускания, f_c – параметр, определяющий ширину полосы фильтра, f_c – центральная частота, ϕ – фаза [2].

При умеренных уровнях мощности сигнала $h_k(t)$ описывается следующей формулой [3]:

$$h_k(t) = 24.7 + 0.108 t \quad (2)$$

В данной работе порядок фильтров α выбран равным 4 ($\alpha = 4$), фаза ϕ равна нулю ($\phi = 0$), параметр β равен 1.019 ($\beta = 1.019$) [5]. Центральные частоты банка фильтров (f_c) выбираются в соответствии с психоакустической шкалой Барков [5].

АЧХ 20-канального банка гамма-тон-фильтров приведена на рисунке 2.

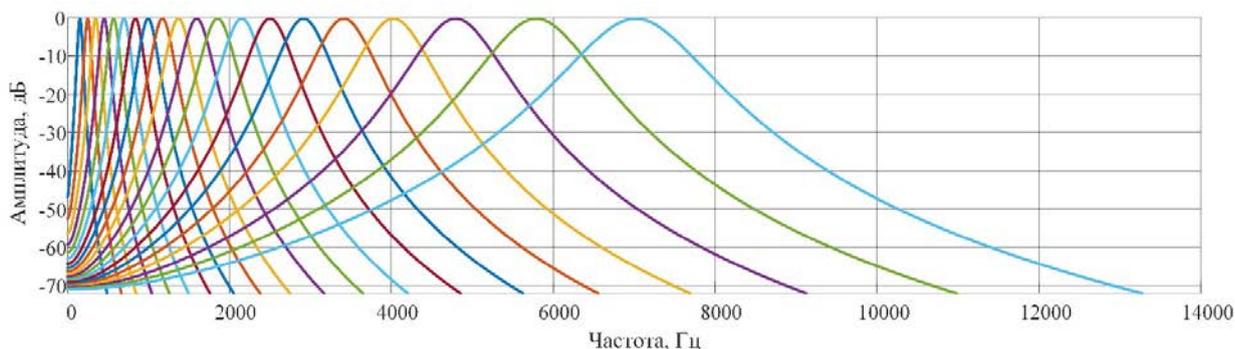


Рисунок 2 – АЧХ банка гамма-тон-фильтров

Рисунок 2 демонстрирует нелинейность восприятия звука человеческим ухом.

Потеря слуха характеризуется сужением слышимого динамического диапазона, снижением порогов слышимости, ухудшением частотного разрешения уха. Следствие этого – разборчивость речи у тугоухого человека падает. Для её повышения необходимо решить задачу отображения широкого динамического диапазона нормально слышащего человека в остаточный диапазон тугоухого человека, т.е. выполнить компрессию динамического диапазона. Помимо этого, необходимо выполнить частотно-зависимое усиление сигнала. Основная цель КДД в разработанной системе – автоматический выбор коэффициентов усиления субполос в зависимости от уровня мощности сигнала. Для выполнения задачи частотно-зависимого усиления сигнала для каждой субполосе строится отдельный компрессор, настраиваемый согласно характеристикам слуха (аудиограмме) тугоухого человека. Использование такого механизма настройки системы позволяет ослаблять субполосы с высокой мощностью, а с малой – усиливать. Благодаря этому, тихие звуки становятся слышимыми, а громкие звуки не вызывают дискомфортных ощущений [6].

Порог компрессии – параметр КДД, определяющий точку изгиба характеристики вход/выход компрессора, после которой алгоритм компрессии становится активным. Если уровень входного сигнала ниже порога компрессии, то выходной сигнал усиливается линейно, а если выше – коэффициент усиления уменьшается [5]. Пример характеристики вход/выход КДД для частоты 153 Гц, построенной по аудиограмме из рисунка 4, приведён на рисунке 3.

Настройка системы под характеристики слуха конкретного тугоухого человека осуществляется на центральных частотах банка фильтров (f_c) с использованием аудиограммы. Аудиограмма представляет собой таблицу минимальных уровней восприятия тональных сигналов на

фиксированном наборе частот и формируется в результате аудиометрии. Проведя анализ аудиограммы можно диагностировать нарушения и тип патологий слуха. На бланке аудиограммы отображаются две оси: горизонтальная — частота в Герцах и вертикальная — минимальные уровни звукового давления сигнала, при которых человек слышит звук (порог слышимости) [5]. Пример аудиограммы для потери слуха второй степени приведён на рисунке 4.

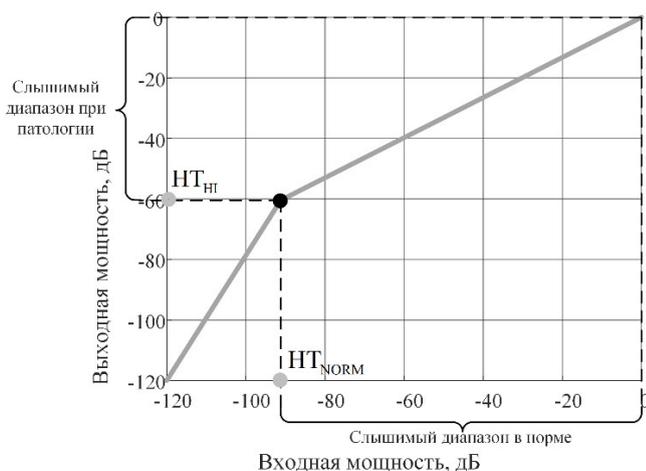


Рисунок 3 – Пример характеристики вход/выход КДД

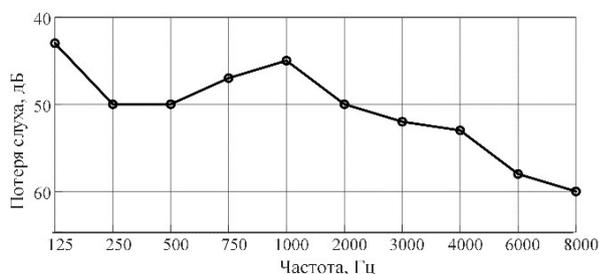


Рисунок 4 – Пример аудиограммы

Помимо характеристики вход/выход КДД, рисунок 3 демонстрирует изменение слышимого диапазона при наличии потери слуха. Так, например, слышимый диапазон для слуха в норме на рисунке 3 составляет 0...-90 дБ, а для тугоухого человека: 0...-60 дБ, что соответствует эффекту уменьшения слышимого динамического диапазона при ухудшении слуха.

Для настройки системы необходимо знать пороги слышимости в норме и при тугоухости на частотах. Аудиограмма используется для задания характеристики вход/выход: из неё определяется исходная кривая, исходя из стандарта ISO226, описывающего кривые слышимости. Зная пороги слышимости в норме и при наличии патологии можно найти точку перегиба характеристики вход/выход КДД (рисунок 5). Для этого, на первом этапе выполняется интерполяция порогов слышимости (в норме и с патологией) на сетку центральных частот банка фильтров. Далее, необходимо выполнить перевод значений порогов слышимости из УЗД (уровень звукового давления) в «цифровые» дБ. Следующим этапом определяется точка перегиба характеристики вход/выход КДД, где в качестве входного используется значение порога слышимости в норме, а выходного – значение порога слышимости тугоухого человека [6, 7].



Рисунок 5 – Блок-схема алгоритма настройки КДД

Для реализации КДД в каждой субполосе на выходе банка фильтров, на частотах производится оценка мощности сигнала:

$$P_i[k] = \begin{cases} P_i[k-1] + \alpha_i^2 P_i[k] (1 - \alpha_i), & \text{если } P_i[k] > P_i[k-1], \\ P_i[k-1] + \alpha_i^2 P_i[k] (1 - \alpha_i), & \text{если } P_i[k] \leq P_i[k-1], \end{cases} \quad \alpha_i = 1, 2, \dots, N_i \quad (3)$$

где $P_i[k]$ – мощность сигнала на выходе i -го канала банка фильтров, N_i – число каналов банка фильтров, α_i – сигнал на выходе i -го канала банка фильтров, α_i – коэффициент атаки, α_i – коэффициент восстановления, которые вычисляются по следующим формулам [6]:

$$\alpha_i = 1 - \frac{1}{10^{0.23 \cdot T_i}} \quad (4)$$

$$\alpha_i = 1 - \frac{1}{10^{0.23 \cdot T_i}} \quad (5)$$

где T_i – период дискретизации, T_i – время атаки в секундах, T_i – время восстановления в секундах [6].

В данной работе параметры α_i приняты равными 0.99 и 0.9992 ($\alpha_i = 0.99$, $\alpha_i = 0.9992$).

Далее, рассчитываются субполосные коэффициенты усиления G_i .

1) полученный уровень мощности P_i переводится в дБ:

$$G_i = 10 \log_{10} \left(\frac{P_i}{P_{ref}} \right)$$

- 2) с использованием линейной интерполяции рассчитывается выходная мощность $P_{out}^{(i)}$ исходя из характеристики вход/выход КДД;
- 3) вычисляются коэффициенты усиления в каждой субполосе (в дБ):

$$G_{i,d} = P_{out}^{(i)} - P_{in}^{(i)} \quad (5)$$

$$G_{i,d} = G_{i,d} + G_{i,d}$$

- 4) коэффициенты $G_{i,d}$ переводятся из дБ в линейный масштаб:

$$g_{i,d} = 10^{G_{i,d}/20} \quad (6)$$

[8]: После нахождения коэффициентов $g_{i,d}$ они умножаются на соответствующие субполосы $S_i[f]$

$$S_{out}[f] = S_{in}[f] \cdot g_{i,d} \quad (7)$$

Выходной сигнал синтезируется путём суммирования субполосных компонент сигнала.

Для проверки работоспособности системы выполнено её MATLAB-моделирование. В качестве входного сигнала использовался речевой сигнал, временное представление и спектрограмма которого представлены на рисунке 6. Настройка системы выполнена согласно аудиограмме, показанной на рисунке 5. Спектрограмма выходного сигнала и его временное представление приведены на рисунке 7.

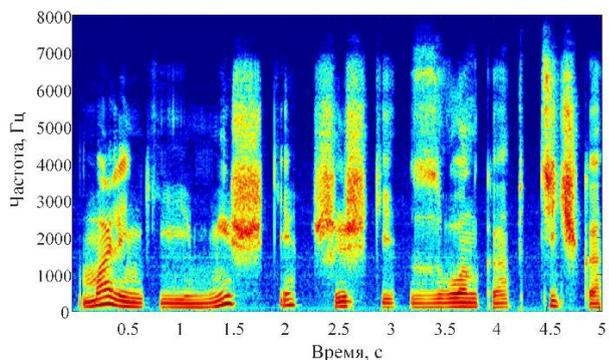


Рисунок 6 – Спектрограмма и временное представление входного сигнала

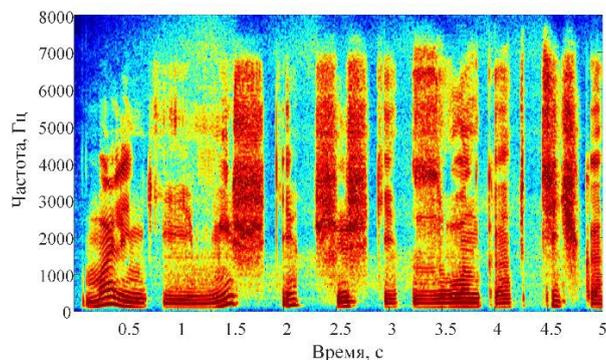


Рисунок 7 – Спектрограмма и временное представление выходного сигнала

Из полученных спектрограмм видно, что используемые в системе КДД выполняют частотно-зависимое усиление, что, в свою очередь, приводит к увеличению уровня исходного звукового сигнала и положительно влияет на разборчивость речи. Искажения, появляющиеся в результате обработки не влияют на общий уровень разборчивости речи и возможность идентифицировать говорящего человека.

Таким образом, в работе рассмотрена система повышения разборчивости речи на основе модели слухового восприятия и компрессии динамического диапазона сигнала. Важной особенностью системы является применение в ней банка гамматон-фильтров, имитирующего механизм частотной декомпозиции звука в ухе. Другой особенностью системы является её автоматическая настройка по аудиограмме слабослышащего человека. Работоспособность системы показана путём MATLAB-моделирования.

Список использованных источников:

1. Всемирная организация здравоохранения [Электронный ресурс] / Глухота и потеря слуха. - Режим доступа: <https://www.who.int/ru/news-room/fact-sheets/detail/deafness-and-hearing-loss>. - Дата доступа: 08.04.2021.
2. R. D. Patterson, I. Nimmo-Smith, J. Holdsworth, P. Rice An efficient auditory filterbank based on the gammatone function / A meeting of the IOC Speech Group on Auditory Modelling at RSRE. 2 (7) – 1987.
3. E. Ambikairajah, J. Epps, L. Lin Wideband speech and audio coding using gammatone filter banks / School of Electrical Engineering and Telecommunications, – 2021.
4. Y. Jiang, Y. Zu, L. Liu, Q. Wang, P. Ren and H. Zhou Gammatone filterbank based energy masking algorithm for active hearing protection system / 5th International Conference on BioMedical Engineering and Informatics (BMEI 2012), Chongqing, China, – 2012. – P. 537–540.
5. U. Zolzer Digital audio signal processing / John Wiley & Sons Software, – 2008. –334 p.
6. М.И. Вашкевич, И.С. Азаров, А.А. Петровский, Косинусно-модулированные банки фильтров с фазовым преобразованием: реализация и применение в слуховых аппаратах // Вашкевич М.И., Азаров И.С., Петровский А.А. – Москва: Горячая линия «Телеком», 2014 – 219 с.: ил.
7. Вашкевич, М. И. Бионспирированный и психоакустически мотивированный анализ звуковых сигналов / Вашкевич М. И., Азаров И. С. // Информационные технологии и системы 2020 (ИТС 2020) = Information Technologies and Systems 2020 (ITS 2020) : материалы международной научной конференции, Минск, 18 ноября 2020 г. / Белорусский государственный университет информатики и радиоэлектроники ; редкол. : Л. Ю. Шилин [и др.]. – Минск, 2020. – С. 9–12.
8. М. И. Вашкевич, А. А. Петровский Компрессор речевых сигналов для слуховых аппаратов на основе банка кохлеарных фильтров / Цифровая обработка сигналов и ее применение: труды 14-й междунар. конф., Россия, Москва. – 2012. – Т. 1. – С. 87–91.

UDC 534.773

SYSTEM FOR IMPROVING SPEECH INTELLIGIBILITY BASED ON AN AUDITORY PERCEPTION MODEL AND DYNAMIC SIGNAL RANGE COMPRESSION

Semenuk N.A.

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus

Porhun M.I. – Master of Engineering Science

Annotation. The paper considers a system for improving speech intelligibility based on an auditory perception model and dynamic signal range compression. The functioning of the system is reduced to sub-band processing of the audio signal coordinated with human auditory system mechanism. A bank of gammaton-filters is used to decompose an audio signal into components (sub-bands). Each subband passes through its own dynamic range compressor whose task is to automatically select the gain for the current subband. The amplified sub-bands are synthesized into the output signal by summation. The system configuration is based on the audiogram of a deaf person. A MATLAB simulation was performed to check the system's performance.

Keywords. Speech intelligibility enhancement, gammaton-filter, dynamic range compressor, audiogram.