

МЕТОД ПЕРЕНОСА ЧАСТОТ В РЕЧЕВОМ СИГНАЛЕ ДЛЯ КОРРЕКЦИИ СЛУХОВЫХ ПАТОЛОГИЙ

магистр. Порхун М.И., доц. Вашкевич М.И.

Белорусский государственный университет информатики и радиоэлектроники
ул. П. Бровки, 6, БГУИР, каф. ЭВС, 220013, Минск, Беларусь, e-mail: maxim.porhun@mail.ru,
vashkevich@bsuir.by

Аннотация. Предлагается метод переноса частот в речевом сигнале для коррекции слуховых патологий. Задача переноса частот сводится к обработке сигнала таким образом, чтобы тугоухий человек смог воспринимать высокочастотные компоненты окружающей акустической среды. Метод заключается в переносе высокочастотной составляющей спектра сигнала в низкочастотную область без компрессии. Размеры диапазонов переносимых составляющих спектра сигнала являются фиксированными. В работе приводится пример переноса фиксированного диапазона частот в речевом сигнале без компрессии.

Введение. У большинства людей, страдающих от потери слуха, сохраняется способность к восприятию звука в низкочастотной области. Однако высокочастотные составляющие, в большей степени, они уже не слышат. Это приводит к тому, что для тугоухого человека затруднено восприятие многих согласных звуков, энергия спектра которых сосредоточена в высокочастотной области. Замечено, что у таких людей резко ухудшается разборчивость речи при наличии фонового шума. В зависимости от степени потери слуха, простое усиление частотных компонентов сигнала зачастую оказывается не эффективным при попытке улучшить разборчивость высокочастотных компонентов звука. Этот факт объясняется тем, что часть волосковых клеток в улитке уха, отвечающих за восприятие высокочастотных звуков, мертвы. Как следствие этого, для людей с тяжёлой степенью потери слуха, невозможно восстановить разборчивость высокочастотных звуков только при помощи усиления [1]. Однако, решением данной задачи может послужить перенос высокочастотных компонентов звука в низкочастотную область, в которой тугоухий человек сохраняет возможность восприятия.

Метод переноса частот. Для переноса высокочастотных компонентов сигнала в низкочастотную область предлагается использовать схему, приведённую на рисунке 1.

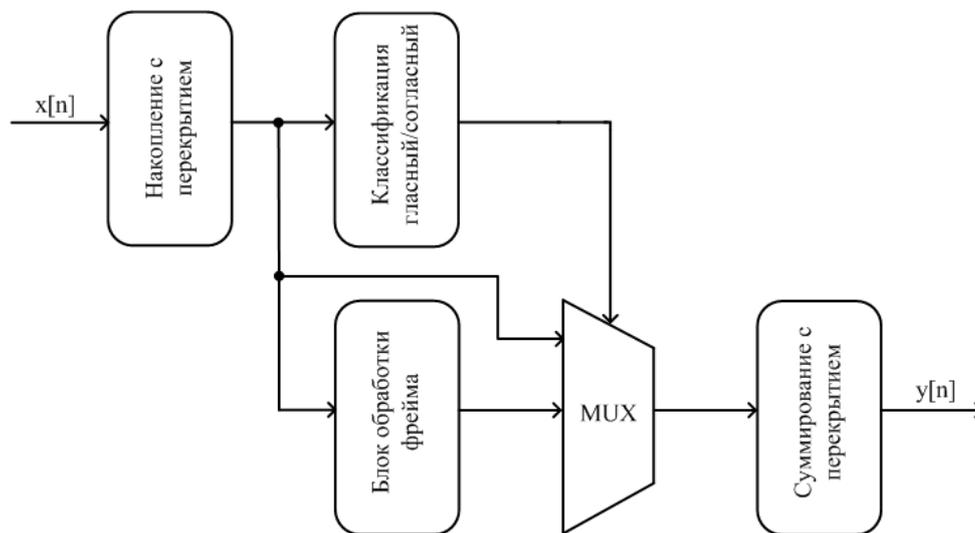


Рисунок 1 – Схема переноса частот в речевом сигнале для коррекции слуховых патологий

Входной сигнал $x[n]$ разделяется на перекрывающиеся фреймы. Каждый фрейм классифицируется по признаку гласный или согласный звук в нём содержится. Фреймы, содержащие согласные звуки, поступают на выход после соответствующей обработки, а фреймы, содержащие гласные звуки оставляются без изменений. В блоке обработки фреймов реализован механизм переноса частот спектра сигнала. Обработанные фреймы синтезируются в выходной сигнал $y[n]$ путём суммирования с перекрытием.

Далее в работе описываются основные аспекты метода переноса частот в речевом сигнале для коррекции слуховых патологий.

Блок обработки фрейма. Для обработки каждого фрейма предлагается использовать схему, приведённую на рисунке 2.

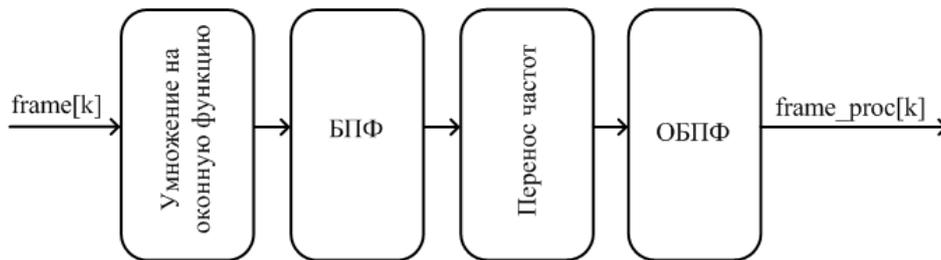


Рисунок 2 – Общая структура блока обработки фрейма

Для сглаживания уровня сигнала на концах фрейма, каждый фрейм умножается на временное окно и переводится в частотную область при помощи быстрого преобразования Фурье (БПФ). Далее происходит непосредственно сам перенос высоких частот в низкочастотную область. Полученный фрейм переводится обратно во временную область при помощи обратного БПФ (ОБПФ) [2].

Классификация фрейма. Согласные звуки находятся в высокочастотной части спектра речевого сигнала, а гласные – в низкочастотной. Именно поэтому тугоухие люди хуже воспринимают согласные звуки, чем гласные [3]. Для классификации звуков в работе использовался классификатор, состоящий из четырёх параметров:

1) среднее число переходов через нуль [4]:

$$Z = \frac{1}{N} \sum_{n=1}^N |\text{sgn}(x[n]) - \text{sgn}(x[n-1])| \quad (1)$$

2) кратковременная энергия [5]:

$$STE = \sum_{n=1}^N x^2[n] \quad (2)$$

3) кратковременная амплитуда [5]:

$$STM = \sum_{n=1}^N |x[n]| \quad (3)$$

4) нормализованная автокорреляция [6]:

$$NC = \frac{\sum_{n=1}^N x[n] x[n-1]}{\sqrt{\left(\sum_{n=1}^N x^2[n]\right) \left(\sum_{n=1}^N x^2[n-1]\right)}} \quad (4)$$

где $x[n]$ – значение текущего отсчёта сигнала, N – количество отсчётов в рассматриваемом сигнале, а

$$\text{sgn}(x[n]) = \begin{cases} 1, & \text{если } x[n] \geq 0 \\ -1, & \text{если } x[n] < 0. \end{cases} \quad (5)$$

Для определения целевых значений вышеперечисленных параметров для согласных и гласных звуков, а также пауз в речи было произведено моделирование на тестовом речевом сигнале длительность 30 секунд. Тестовый сигнал был размечен по временным промежуткам,

соответствующим гласным и согласным звукам, а также паузам. По результатам моделирования была получена таблица значений целевых параметров, содержащая три класса:

- 1) гласный звук;
- 2) согласный звук;
- 3) пауза.

Для определения того, содержит фрейм гласный или согласный звук, или же паузу для каждого фрейма вычислялись параметры 1-4. При помощи метода «ближайшего соседа» (K-Nearest Neighbor) [7] и полученных для текущего фрейма значений целевых параметров, а также таблицы с заранее вычисленными параметрами для классификации звуков выбирается наиболее подходящий для текущего фрейма класс.

Механизм переноса частот. Для переноса частот выбираются две частотных полосы: исходная и целевая. Выбор полос должен происходить весьма тщательно. Если целевая полоса слишком низкая, то можно повредить низкочастотный диапазон и негативно повлиять на остаточный слух в диапазоне низких частот. Если же выбрать целевую полосу слишком высокочастотной, то перенос частот не будет иметь никакого смысла, так как этот механизм не будет сохранять высокочастотную информацию на достаточно низких частотах, которые слышат люди с остаточным слухом. В работе выбрана целевая полоса 750 – 1750 Гц и исходная полоса – 2000 – 3000 Гц. Целевая полоса выбрана исходя из того, чтобы сохранить большую часть гласных звуков и естественную высоту звука, поскольку частоты ниже 750 Гц остаются неизменными. Фиксированная исходная полоса может не в полной мере представлять различные речевые сигналы с различными частотными компонентами. Для компенсации этого недостатка предлагается использовать исходную адаптивную частотную полосу. Первый спектральный момент, представляющий центроид спектра, хорошо характеризует согласные звуки. Поэтому при обработке фреймов, предлагается использовать исходную адаптивную частотную полосу с центральной частотой M , которая рассчитывается следующим образом:

$$M = \frac{\sum_{l=1}^{N/2} P(l) (f_s * \frac{1}{N})}{\sum_{l=1}^{N/2} P(l)} \quad (6)$$

где $P(l)$ – мощность l -го частотного отсчёта, f_s – частота дискретизации и N – размер БПФ.

В работе не используется компрессия частотных полос, то есть целевая и исходная частотные полосы имеют одинаковую ширину [2].

Для сглаживания границ исходной полосы с центральной частотой M предлагается использовать экспоненциальное усреднение:

$$low_border[n] = \alpha \times low_border + (1 - \alpha) \times low_border[n - 1] \quad (7)$$

где $low_border[n]$ – нижняя граница исходной полосы для текущего фрейма, $low_border[n - 1]$ – нижняя граница исходной полосы для предыдущего фрейма, коэффициент α выбран равным 0.995.

Поскольку диапазон исходной полосы фиксирован, то усредняется только нижняя граница относительно центроида спектра M . Верхняя граница вычисляется относительно нижней.

Результаты экспериментов. Проверка предлагаемого метода переноса частот для коррекции слуховых патологий производилась на речевом сигнале (рисунок 3, в). Спектрограмма исходного речевого сигнала приведена на рисунке 3, а. Спектрограмма (рисунок 3, б) выходного сигнала (рисунок 3, г) демонстрирует механизм переноса частот.

В качестве примера в работе используется входной речевой сигнал $x[n]$ с частотой дискретизации 16 кГц. Он разделяется на фреймы длительностью 16 мс и умножаются на окно Хемминга. Целевой и исходной частотными полосами выбраны 750 – 1750 Гц и 2000 – 3000 Гц соответственно. На рисунке 3,б выделен пример переносимых диапазонов частот.

Дальнейшие исследования. В дальнейших исследованиях планируется значительно улучшить механизм переноса частот. Для улучшения классификации фреймов по признаку гласный или согласный звук планируется использовать несколько признаков (оценка энергии спектра фрейма, нормализованной автокорреляции и амплитуды) [5,6]. Вместо фиксированной

целевой и адаптивной исходной частотных полос планируется использовать настраиваемые по аудиограмме тугоухого человека частотные полосы. Также планируется использовать механизм настраиваемого взвешенного переноса частот, который будет учитывать особенности восприятия окружающей акустической среды конкретным человеком [2].

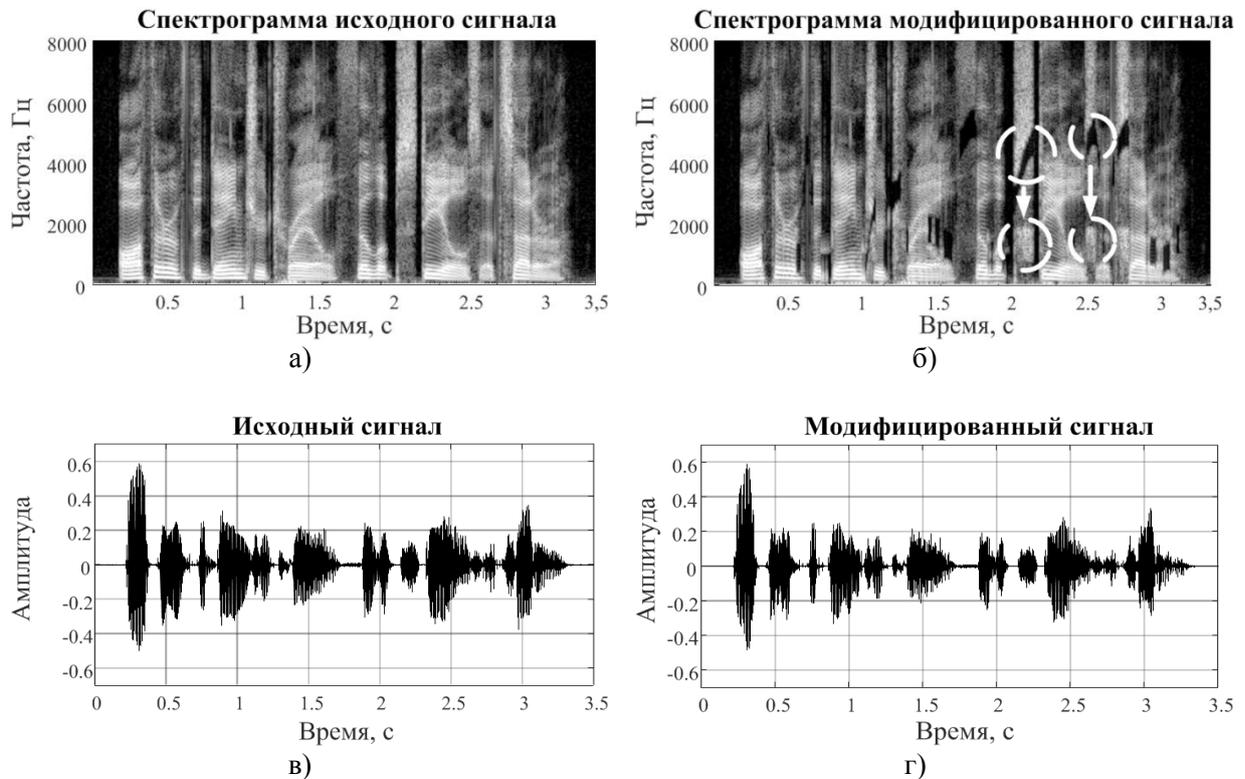


Рисунок 3 – Результаты обработки сигнала предложенным методом: а) спектрограмма исходного сигнала; б) спектрограмма модифицированного сигнала; в) исходный сигнал; г) модифицированный сигнал

Вывод. В работе предложен метод переноса частот для коррекции слуховых патологий. Особенностью метода является относительная простота его реализации и большая перспектива развития. Данный метод может быть использован в методах коррекции слуха. Работоспособность метода подтверждена результатами MATLAB-моделирования.

Литература

- [1] A. Simpson Frequency-lowering devices for managing high-frequency hearing loss: A review / Trends in Amplification, – 2009. – Vol. 13, No. 2. – P. 87 – 106.
- [2] Y. Liu, R. Y. Chang, Y. Tsao and Y. Chang A new frequency lowering technique for Mandarin-speaking hearing aid users / IEEE Global Conference on Signal and Information Processing (GlobalSIP), Orlando, FL, – 2015. P. 722 – 726.
- [3] C. A. Hogan and C. W. Turner High-frequency audibility: Benefits for hearing-impaired listeners / The Journal of the Acoustical Society of America, – 1998. – Vol. 104, No. 1. – P. 432 – 441.
- [4] Л.Р. Рабинер, Р.В. Шафер Цифровая обработка речевых сигналов / Радио и связь, Москва, – 1981. – 496 С.
- [5] M. Jalil, F. A. Butt and A. Malik Short-time energy, magnitude, zero crossing rate and autocorrelation measurement for discriminating voiced and unvoiced segments of speech signals / The International Conference on Technological Advances in Electrical, Electronics and Computer Engineering (TAECE), Konya, – 2013. – P. 208 – 212.

[6] Y.-T. Liu, Y. Tsao, and R. Y. Chang A deep neural network based approach to Mandarin consonant/vowel separation / IEEE International Conference on Consumer Electronics (ICCE- TW), Taiwan, – 2015. – P. 324 – 325.

[7] Ranny Voice recognition using k Nearest Neighbor and double distance method / International Conference on Industrial Engineering, Management Science and Application (ICIMSA), Jeju, – 2016. – P. 1 –5.

FREQUENCY TRANSPOSITION METHOD IN A SPEECH SIGNAL FOR HEARING LOSS CORRECTION

Porhun M.I., Vashkevich M.I.

Belarusian State University of Informatics and Radioelectronics
6, P. Brovki str., Computer Engineering Department, 220013, Minsk, Belarus,
e-mail: maxim.porhun@mail.ru, vashkevich@bsuir.by

Abstract. Frequency transposition method in a speech signal for hearing loss correction is proposed. Frequency transposition aims at the signal processing in such a way that the deaf-eared person can perceive the surrounding high-frequency components of the acoustic environment. The method based on the high-frequency signal components transposition to the low-frequency region of the spectrum without compression. The sizes of the transposed ranges of the signal spectrum are fixed. Frequency transposition example in a speech signal of the fixed bandwidth is presented in the article.