



<http://dx.doi.org/10.35596/1729-7648-2020-18-1-43-51>

Оригинальная статья
Original paper

УДК 534.773

МЕТОД КОРРЕКЦИИ СЛУХА НА ОСНОВЕ ПСИХОАКУСТИЧЕСКИ ОБУСЛОВЛЕННОГО ПЕРЕНОСА ЧАСТОТ В РЕЧЕВОМ СИГНАЛЕ

ПОРХУН М.И., ВАШКЕВИЧ М.И.

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

Поступила в редакцию 22 июля 2019

© Белорусский государственный университет информатики и радиоэлектроники, 2020

Аннотация. Целью работы являлась разработка метода обработки речевого сигнала для коррекции слуховых патологий на основе психоакустически обусловленного переноса высокочастотных составляющих спектра сигнала в низкочастотную область с последующим частотно-зависимым усилением. Для достижения поставленной цели были решены задачи, связанные с разработкой принципов переноса частот в речевом сигнале. Разработанный метод является адаптивным, его настройка осуществляется согласно аудиограмме тугоухого человека. Для переноса частот выбираются две частотных полосы: исходная (откуда производится перенос) и целевая (куда производится перенос). Ширина исходной частотной полосы фиксирована, а ширина целевой полосы выбирается адаптивно. Перенос спектра выполняется только для согласных звуков, восприятие которых тугоухими людьми затруднено. Классификация звуков по признаку гласный/согласный/пауза реализована на базе нейронной сети. В качестве информационных признаков выбирались: среднее число переходов через нуль, кратковременная энергия, кратковременная амплитуда, нормализованная автокорреляционная функция и первый спектральный момент. Чтобы сохранить максимально натуральное звучание переносимых звуков используется концепция равной громкости. Для компенсации ослабления восприятия звука тугоухим человеком используется частотно-зависимое усиление сигнала на основе аудиограммы. Эффективность предложенного метода проверена экспериментально с использованием моделирования эффекта потери слуха. В эксперименте участвовали 10 человек, которым давали прослушивать записи, пропущенные через модель потери слуха, а также записи, пропущенные через модель потери слуха с последующей коррекцией. Результаты показали, что применение разработанного метода коррекции слуха в среднем улучшает разборчивость речи на 6 %.

Ключевые слова: коррекция слуха, слуховые патологии, моделирование потери слуха.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Для цитирования. Порхун М.И., Вашкевич М.И. Метод коррекции слуха на основе психоакустически обусловленного переноса частот в речевом сигнале. Доклады БГУИР. 2020; 18(1): 43-51.

HEARING CORRECTION METHOD BASED ON PSYCHOACOUSTICALLY MOTIVATED FREQUENCY TRANSPOSITION IN A SPEECH SIGNAL

MAXIM.I. PORHUN, MAXIM.I. VASHKEVICH

*Belarusian state university of informatics and radioelectronics
P. Brovki str., 6, Minsk, 220013, Republic of Belarus*

Submitted 22 July 2019

© Belarusian State University of Informatics and Radioelectronics, 2020

Abstract. The purpose of the work was to develop a speech signal processing method for the hearing pathologies correction based on psychoacoustically motivated transposition of high-frequency components of the signal spectrum to the low-frequency region with subsequent frequency-dependent amplification. To achieve this goal, several tasks related to the development of principles of frequency transposition in a speech signal were solved. The adjustment of the method is carried out according to the audiogram of a deaf person. For frequency transposition, source and target frequency bands are selected. The width of the source frequency band is fixed, while the width of the target band is adaptive. Spectrum transposition is performed only for consonants, the perception of which is more difficult for people with hearing loss. The classification of sounds (into vowel-consonant - pause classes) is implemented using one-layer neural network. The feature vector consists of: the zero crossing rate, short-term energy, short-term magnitude, normalized autocorrelation function and the first spectral moment. To preserve the naturalness of transposed sounds, the concept of equal loudness is used. To compensate for the attenuation in the perception of sound by a deaf person, a frequency-dependent signal amplification based on an audiogram is used. The effectiveness of the proposed method was verified experimentally using hearing loss effect simulation. The experiment involved 10 people who were given to listen to the recordings passed through the hearing loss model, as well as recordings passed through the hearing loss model with subsequent correction (using proposed method). The results showed that using the proposed hearing correction method improves speech intelligibility on average by 6 %.

Keywords: hearing correction, hearing impairments, hearing loss simulation.

Conflict of interests. The authors declare no conflict of interests.

For citation. Porhun M.I., Vashkevich M.I. Hearing correction method based on psychoacoustically motivated frequency transposition in a speech signal. Doklady BGUIR. 2020; 18(1): 43-51.

Введение

По данным Всемирной организации здравоохранения, в мире происходит рост числа людей, страдающих от патологий слуха, что стимулирует разработку различных методов коррекции слуха. У большинства слабослышащих затруднено восприятие высокочастотных (ВЧ) звуков, при этом чувствительность к низкочастотным (НЧ) звукам, как правило, сохраняется. В зависимости от типа и степени тугоухости применение классических методов коррекции слуха не всегда приводит к повышению разборчивости речи [1]. Для людей с тяжелыми слуховыми патологиями и вовсе невозможно восстановить разборчивость звуков, используя только стратегию усиления, поскольку у них мертва часть волосковых клеток в ухе, отвечающих за восприятие ВЧ звуков [1–3]. Решением данной проблемы может служить метод коррекции слуха, основанный на переносе ВЧ части спектра сигнала, не воспринимаемой тугоухим человеком, в НЧ область, где он сохраняет возможность слышать.

Метод коррекции слуха на основе переноса частот

Для коррекции слуховых патологий предлагается метод, который схематично представлен на рис. 1. Входной сигнал поступает на блок переноса частот,

где из высокочастотной области в низкочастотную переносится часть спектра сигнала. Далее выполняется частотно-зависимое усиление сигнала с использованием фильтра, амплитудно-частотная характеристика которого рассчитывается исходя из аудиограммы. Для компенсации эффекта «клиппинг» используется ограничитель уровня сигнала (лимитер). Далее сигнал проходит через фильтр нижних частот (ФНЧ) с частотой среза, равной верхней границе целевой полосы, используемой при переносе частот. Последний шаг необходим для сосредоточения основной энергии сигнала в области, где тугоухий человек сохраняет возможность слышать. В результате обработки синтезируется выходной сигнал, предоставляющий тугоухому человеку возможность слышать ранее не различимые звуки.

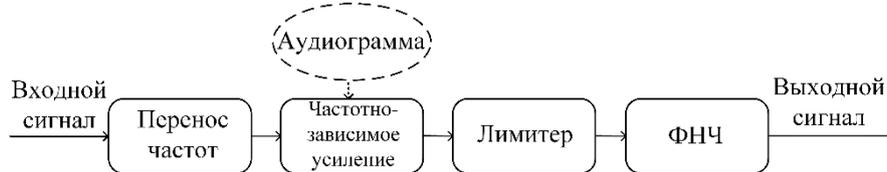


Рис. 1. Этапы обработки сигнала в методе коррекции слуха
Fig. 1. Signal processing stages in hearing correction method

Перенос частот в речевом сигнале

В большинстве случаев согласные звуки занимают высокочастотный диапазон спектра речевого сигнала, а гласные – низкочастотный. Именно поэтому тугоухие люди хуже воспринимают согласные звуки, чем гласные [4]. Для переноса ВЧ компонентов сигнала в НЧ область предлагается использовать схему, приведенную на рис. 2.

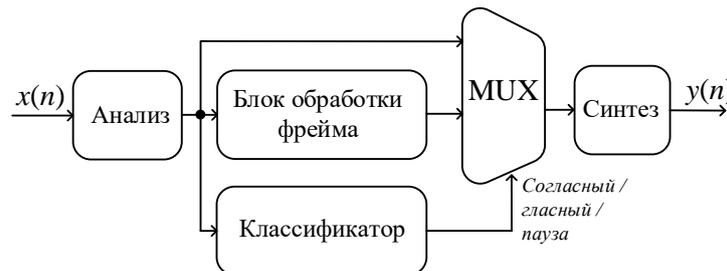


Рис. 2. Схема блока переноса частот
Fig. 2. Frequency transposition block scheme

Входной сигнал разделяется на перекрывающиеся фреймы. Каждый фрейм классифицируется как содержащий гласный/согласный звук или паузу. Фреймы, содержащие согласные звуки, обрабатываются в блоке обработки фрейма, а остальные фреймы оставляются без изменений. Синтез выходного сигнала выполняется методом суммирования с перекрытием. Схема обработки отдельного фрейма приведена на рис. 3.

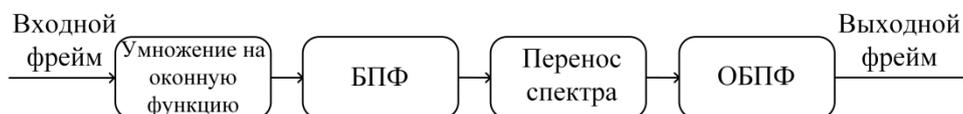


Рис. 3. Структура блока обработки фрейма
Fig. 3. Frame processing block structure

Для сглаживания уровня сигнала на концах фрейма поступающие отсчеты сигнала умножаются на оконную функцию Хэмминга и переводятся в частотную область при помощи быстрого преобразования Фурье (БПФ). Далее происходит перенос высоких частот в НЧ область. Полученный фрейм переводится обратно во временную область обратным БПФ (ОБПФ).

Поскольку слуховая система человека имеет разную чувствительность к звукам разной частоты [5], то прямой перенос спектра приведет к искажению уровня громкости сигнала, и, как следствие, к ухудшению разборчивости речи в целом. Чтобы уменьшить искажения,

предлагается использовать психоакустический принцип равной громкости [6], который позволит сохранить исходную громкость компонент звука при переносе в НЧ область.

Метод переноса частот (см. рис. 4) подразумевает использование двух частотных полос: исходной (откуда выполняется перенос) и целевой (куда выполняется перенос). Чтобы не повредить звуки в НЧ диапазоне и при этом отобразить ВЧ информацию на достаточно низких частотах, выбрана целевая полоса с адаптивной нижней границей f_T^L , располагающейся не ниже 750 Гц, и верхней границей f_T^U , равной 1750 Гц. Ширина исходной полосы ΔW_s составляет 1 кГц, а ее центральная частота f_c выбирается адаптивно (см. рис. 4). Нижняя граница целевой полосы f_T^L выбирается так, чтобы ширина исходной и целевой полосы в шкале барков совпадали. Такой подход позволяет согласовать психоакустический «масштаб» переносимых компонент в сигнале.

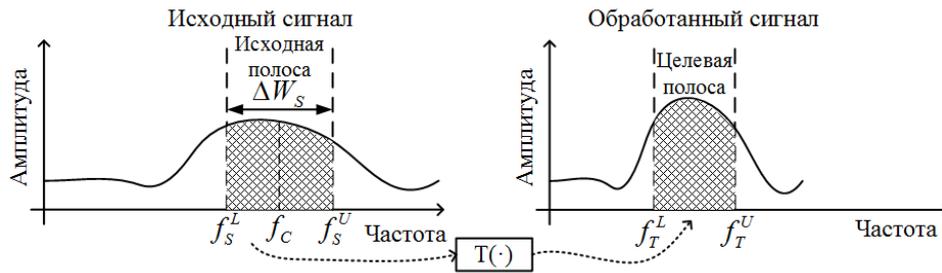


Рис. 4. Общая схема переноса частот
Fig. 4. Frequency transposition general scheme

Первый спектральный момент, представляющий собой центроид спектра, хорошо характеризует согласные звуки. Поэтому для определения исходной полосы предлагается использовать частоту, соответствующую центроиду спектра:

$$f_c = \frac{\sum_{i=1}^{N/2} P(i) \cdot (f_s / N)}{\sum_{i=1}^{N/2} P(i)}, \quad (1)$$

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

Для согласования значений f_c , получаемых на смежных фреймах, предлагается использовать экспоненциальное усреднение:

$$f_c(i) = \alpha \cdot f_c + (1 - \alpha) \cdot f_c(i - 1), \quad (2)$$

где α – сглаживающий коэффициент, выбранный равным 0,995.

Верхняя f_s^U и нижняя f_s^L границы исходной полосы определяются как:

$$f_s^L = f_c - \Delta W_s / 2, \quad f_s^U = f_c + \Delta W_s / 2. \quad (3)$$

Как указывалось ранее, верхняя граница целевой полосы фиксирована ($f_T^U = 1750$ Гц), значит, остается определить только нижнюю границу целевой полосы. Это можно сделать исходя из принципа, что ширина исходной и целевой полосы должны совпадать в шкале барков. Таким образом, определение нижней границы целевой полосы заключается в выполнении следующих двух шагов:

1) определение ширины исходной полосы в шкале барков:

$$\Delta B_s = B(f_s^U) - B(f_s^L), \quad (4)$$

где функция B выполняет переход от частоты в Гц к шкале барков [7]:

$$B(f) = 26,81 \cdot (1 + 1960 / f)^{-1} - 0,53; \quad (5)$$

2) непосредственное вычисление нижней границы целевой полосы:

$$f_T^L = B^{-1}(B(f_T^U) - \Delta B_s), \quad (6)$$

где ΔB_T – ширина целевой полосы в барках ($\Delta B_T = \Delta B_S$).

Компенсация уровня громкости частотных компонент

В предлагаемом методе используются кривые равной громкости для определения корректирующих коэффициентов для частотных компонент, переносимых из ВЧ в НЧ область, чтобы сохранить исходную громкость [8].

Для определения корректирующих коэффициентов предлагается следующий алгоритм:

1) вычислить средние уровни сигнала L_T и L_S для целевой и исходной полосы:

$$L_T = \frac{1}{|T_{ind}|} \sum_{k \in T_{ind}} |X(k)|, \quad L_S = \frac{1}{|S_{ind}|} \sum_{k \in S_{ind}} |X(k)|, \quad (7)$$

где T_{ind} и S_{ind} – множества индексов частотных отсчетов в целевой и исходной полосе соответственно, $X(k)$ – кратковременный спектр сигнала;

2) вычислить центральные частоты F_T и F_S для целевой и исходной полос:

$$F_T = (f_T^U - f_T^L)/2, \quad F_S = (f_S^U - f_S^L)/2. \quad (8)$$

3) определить уровень громкости (в фонах) для исходной полосы $p_S = \psi_{F_S}(L_S)$, где $\psi_F(L)$ – функция, переводящая уровень сигнала L на частоте F в громкость p .

4) определить требуемый уровень сигнала \bar{L}_T для целевой частоты F_T , который соответствует громкости p_S исходной полосы: $\bar{L}_T = \psi_{F_T}^{-1}(p_S)$;

5) вычислить разницу между требуемым и текущим уровнем сигнала в целевой полосе:

$$\Delta L_T = \bar{L}_T - L_T; \quad (9)$$

6) вычислить весовой коэффициент для целевой полосы $G = 10^{\Delta L_T/20}$.

Для компенсации уровня громкости сигнала все частотные отсчеты целевой полосы умножаются на полученный весовой коэффициент G . Иллюстрация процесса компенсации уровня громкости представлена на рис. 5.

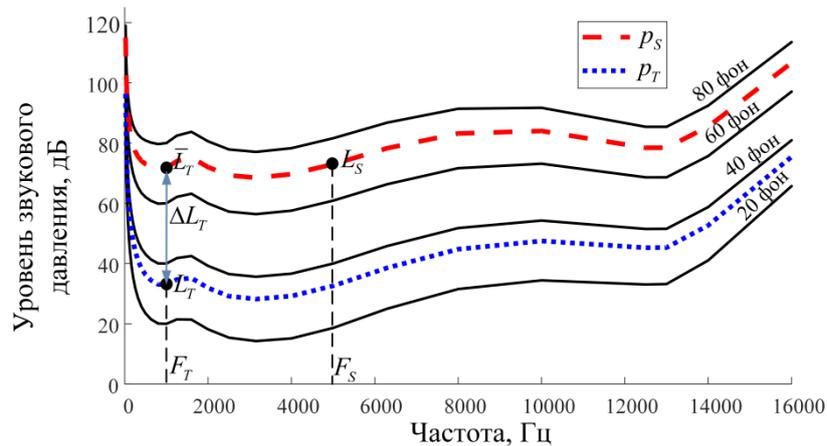


Рис. 5. Общая схема компенсации уровня громкости
Fig. 5. Loud compensation general scheme

Классификация звуков

Классификация звуков необходима для отнесения каждого фрейма к одному из классов: гласный, согласный или пауза. В качестве признаков для классификации использовались следующие параметры: среднее число переходов через нуль (ZCR), кратковременная энергия

(STE), кратковременная амплитуда (STM), нормализованная автокорреляционная функция (NC) и первый спектральный момент (SM) [8].

Для обучения классификатора использовались звуковые файлы с соответствующей разметкой на три класса (гласный/согласный/пауза). На 30 % звуковые файлы состояли из речевой базы «ТИМІТ», а остальные 70 % были записаны и размечены вручную. С целью разделения фреймов по вышеперечисленным классам построена однослойная нейронная сеть с выходной функцией активации «softmax» (рис. 6) [9].

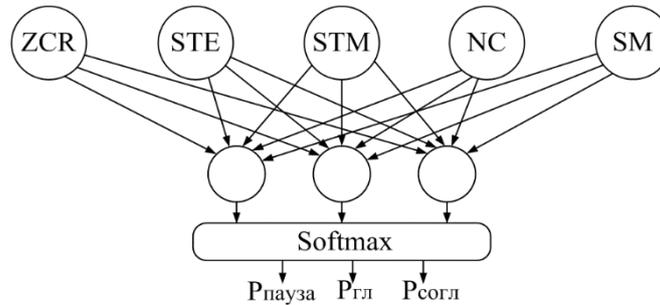


Рис. 6. Структура классификатора
Fig. 6. Classifier structure

80 % звуковых файлов использовались для обучения классификатора, оставшиеся 20 % – для тестирования. В результате тестирования обученного классификатора получена матрица неточностей (см. табл.).

Из табл. видно, что точность распознавания согласных звуков является довольно высокой (93,1 %), что является важным для решаемой задачи. Точность распознавания гласных звуков и пауз несколько ниже (79,5 и 84,1 %, соответственно).

Таблица. Матрица неточностей классификатора звуков
Table. Confusion matrix for classifier of sounds

		Предсказано как Predicted as			Точность, % Accuracy, %
		согласный consonant	гласный vowel	пауза pause	
Фреймы, содержащие Frames containing	согласные consonants	13381	900	90	93.1%
	гласные vowels	342	2138	210	79.5%
	паузы pauses	278	462	3902	84.1%

Частотно-зависимое усиление

Фильтр для коррекции огибающей спектра сигнала является известным методом компенсации ослабления восприятия звука тугоухим человеком. Для расчета фильтра-корректора существуют формулы, выбор которых выполняется самим человеком исходя из субъективных ощущений либо врачом, на основе типа патологии. Для расчета целевого усиления выбрана формула NAL-R [6], на основе которой синтезировался КИХ фильтр-корректор 128 порядка. Для сосредоточения энергии сигнала в полосе, где тугоухий человек сохраняет возможность слышать, предлагается фильтровать часть спектра, расположенную за верхней границей целевой полосы f_T^U , равной 1750 Гц. Фильтрация также необходима для исключения влияния высоких частот при экспериментальном определении эффективности метода коррекции слуха.

Проверка метода коррекции слуха

Проверка предлагаемого метода выполнена на тестовом речевом сигнале в среде MATLAB. Проведен следующий эксперимент: в первом случае тестовый сигнал обработан методом моделирования потери слуха¹, а во втором – методом коррекции слуха и затем методом моделирования потери слуха. В качестве исходных данных для обоих методов использовалась аудиограмма, соответствующая второй степени тугоухости. Результаты обработки тестового сигнала приведены на рис. 7.

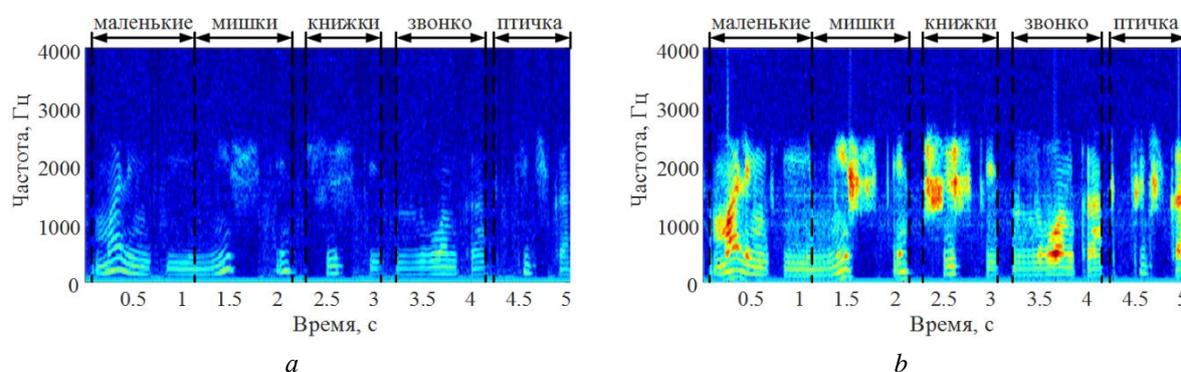


Рис. 7. Результаты экспериментов: *a* – спектрограмма сигнала без обработки методом коррекции слуха; *b* – спектрограмма сигнала, обработанного методом коррекции слуха

Fig. 7. Experimental results: *a* – spectrogram of the signal without processing by hearing correction method; *b* – spectrogram of the signal processed hearing correction method

Из полученных спектрограмм видно, что использование предлагаемого метода коррекции слуха помогает компенсировать уровень громкости сигнала, а также переместить ранее не слышимые звуки (согласные) в слышимую область, что по субъективным ощущениям помогает улучшить разборчивость речи.

Экспериментальные исследования

Одним из способов определения эффективности методов коррекции слуха является оценка посредством выявления процента разборчивости речи при помощи артикуляционных таблиц слов. Процедура состоит в прослушивании испытуемым записанных слов с последующим повторением услышанной информации, правильность которой проверяется [5].

Для тестирования отобрана группа нормально слышащих людей из 10 человек в возрасте от 20 до 66 лет. Для измерения разборчивости речи записаны два тестовых речевых сигнала на основе артикуляционных таблиц из ГОСТ 16600.

Первый тестовый сигнал, содержащий запись 50 слов, пропускаясь через модель потери слуха¹ и отфильтровывался ФНЧ с частотой среза f_T^U , что позволило обеспечить чистоту эксперимента. Второй тестовый сигнал состоял из 50 слов и обрабатывался методом коррекции слуха, а затем пропускаясь через модель потери слуха. Настройка обоих методов осуществлена по аудиограмме, соответствующей второй степени тугоухости. Каждый испытуемый прослушивал оба тестовых сигнала и повторял услышанное слово, корректность которых фиксировалась. График на рис. 8 показывает соотношение разборчивости речи у участников эксперимента с использованием метода коррекции слуха и без него. Результаты эксперимента показали, что применение метода позволило в среднем повысить разборчивость речи на 6 % и максимум – на 18 %.

Поскольку в эксперименте использовалась небольшая выборка участников, то проводилась статистическая проверка значимости результатов при помощи парного *t*-критерия. В результате уровень значимости составил $p = 0,03$, т. е. вероятность того, что

¹ Порхун М.И., Вашкевич М.И. Моделирование потери слуха. *Цифровая обработка сигналов и ее применение: труды 20-й междунар. конф., Россия, Москва, 2018; Т. 1.:228-233.*

отличие среднего уровня разборчивости с применением метода коррекции и без него не являются существенным, очень мала (~3 %).

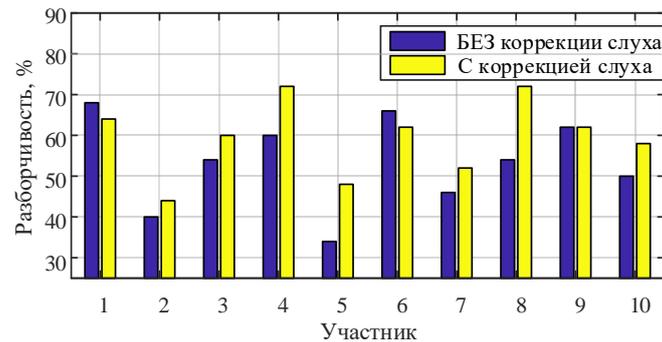


Рис. 8. Результаты эксперимента
Fig. 8. Experimental results

Заключение

В работе предложен метод коррекции слуха, основанный на переносе частотной полосы адаптивной ширины из ВЧ в НЧ область спектра речевого сигнала с последующим частотно-зависимым усилением. Работоспособность метода показана путем MATLAB-моделирования. Эффективность метода доказана экспериментально. Ключевой особенностью предложенного метода является перенос частот в соответствии с психоакустическими аспектами восприятия звука человеком.

Список литературы

1. Simpson A. Frequency-lowering devices for managing high-frequency hearing loss: a review. *Trends in amplification*. 2009;13(2):87-106. DOI: 10.1177/1084713809336421.
2. Alexander J.M. Individual variability in recognition of frequency-lowered speech. *Seminars in Hearing*. 2013;34(2):86-109. DOI: 10.1055/s-0033-1341346.
3. Robinson J.D., Baer T., Moore B. Using transposition to improve consonant discrimination and detection for listeners with severe high-frequency hearing loss. *International Journal of Audiology*. 2007;46(6):293-308. DOI: 10.1080/14992020601188591.
4. Hogan C.A., Turner C.W. High-frequency audibility: Benefits for hearing-impaired listeners. *The Journal of the Acoustical Society of America*. 1998;104:432-441. DOI: 10.1121/1.423247.
5. Королёва И.В. *Введение в аудиологию и слухопротезирование*. СПб: КАРО; 2012.
6. Фонлантен А., Хорст А. *Слуховые аппараты*. Ростов н/Д.: Феникс; 2009.
7. Traunmuller H. Analytical Expressions for the tonotopic sensory scale. *Acoustical Society of America*. 1990; 88(1):97-100. DOI: 10.1121/1.399849.
8. Liu Y.-T., Chang R.Y., Tsao Y., Chang Y.-P. A new frequency lowering technique for Mandarin-speaking hearing aid users. *IEEE Global Conference on Signal and Information Processing (GlobalSIP), Orlando, FL*. 2015;722-726. DOI: 10.1109/GlobalSIP.2015.7418291.
9. Николенко С.И., Архангельская Е.В., Кадурич А.А. *Глубокое обучение. Погружение в мир нейронных сетей*. СПб.: Питер; 2019.

References

1. Simpson A. Frequency-lowering devices for managing high-frequency hearing loss: a review. *Trends in amplification*. 2009;13(2):87-106. DOI: 10.1177/1084713809336421.
2. Alexander J.M. Individual variability in recognition of frequency-lowered speech. *Seminars in Hearing*. 2013;34(2):86-109. DOI: 10.1055/s-0033-1341346.
3. Robinson J.D., Baer T., Moore B. Using transposition to improve consonant discrimination and detection for listeners with severe high-frequency hearing loss. *International Journal of Audiology*. 2007;46(6):293-308. DOI: 10.1080/14992020601188591.

4. Hogan C.A., Turner C.W. High-frequency audibility: Benefits for hearing-impaired listeners. *The Journal of the Acoustical Society of America*. 1998;104:432-441. DOI: 10.1121/1.423247.
5. Korolyova I.V. [Introduction to Audiology and Hearing Prosthetics]. SPb : KARO; 2012. (In Russ.)
6. Vonlanthen A., Horst A. [Hearing Aids]. Rostov n/D: Phoenix; 2009. (In Russ.)
7. Traunmuller H. Analytical Expressions for the tonotopic sensory scale. *Acoustical Society of America*. 1990; 88(1):97-100. DOI: 10.1121/1.399849.
8. Liu Y.-T., Chang R.Y., Tsao Y., Chang Y.-P. A new frequency lowering technique for Mandarin-speaking hearing aid users / *IEEE Global Conference on Signal and Information Processing (GlobalSIP), Orlando, FL*. 2015;722-726. DOI: 10.1109/GlobalSIP.2015.7418291.
9. Nikolenko S.I., Arhangel'skaya E.V., Kadurin A.A. *Glubokoe obuchenie. Pogruzhenie v mir neyronnyih setey*. SPb.: Piter; 2019. (in Russ.)

Вклад авторов

Порхун М.И. разработал метод коррекции слуха на основе психоакустически обусловленного переноса частот в речевом сигнале, а также провел экспериментальные исследования эффективности данного метода.

Вашкевич М.И. определил задачи, которые необходимо было решить в ходе проведения исследований, принимал участие в разработке метода коррекции слуха, участвовал в интерпретации результатов эксперимента.

Authors contribution

Porhun M.I. developed hearing correction method based on psychoacoustically motivated frequency transposition in a speech signal, and also conducted an experimental study of the effectiveness of this method.

Vashkevich M.I. identified the tasks that needed to be solved during the research, and also took part in the development of a hearing correction method, participated in the interpretation of the results of the experiment.

Сведения об авторах

Порхун М.И., ассистент кафедры электронных вычислительных средств Белорусского государственного университета информатики и радиоэлектроники.

Вашкевич М.И., к.т.н., доцент, доцент кафедры электронных вычислительных средств Белорусского государственного университета информатики и радиоэлектроники.

Information about the authors

Porhun M.I., Assistant Lecturer of Computer Engineering Department of Belarusian State University of Informatics and Radioelectronics.

Vashkevich M.I., PhD, Associate Professor, Associate Professor of Computer Engineering Department of Belarusian State University of Informatics and Radioelectronics.

Адрес для корреспонденции

220013, Республика Беларусь,
г. Минск, ул. П. Бровки, д. 6,
Белорусский государственный университет
информатики и радиоэлектроники
тел. +375-17-293-84-20;
e-mail: porhun@bsuir.by
Порхун Максим Игоревич

Address for correspondence

220013, Republic of Belarus,
Minsk, P. Brovki str., 6,
Belarusian State University
of Informatics and Radioelectronics
tel. +375-17-293-84-20;
e-mail: porhun@bsuir.by
Porhun Maxim Igorevich