

nels requires the introduction of so-called cordon bands and it does not allow to realize the potentially achievable transmission rate.

Thus, the development of a method of synthesis of signals of finite duration, optimal in the sense of maximum concentration of energy in specified frequency ranges, is the actual problem. It is in this formulation, the problem of forming of channel signals and is considered in this work.

The method consists in the formation of the channel signal on the basis of eigenvectors with coefficients, which are information bits of the original signal. Eigenvectors are formed on the basis substrip matrix calculated for a given frequency range. The sequence of bits must have a bipolar appearance. This form of the original sequence eliminates the possibility of losing the eigenvector with the multiplication of a zero coefficient. The coefficients can be any number of which increases the speed of the transmitted information.

The probability of correct reception of the transmission of information is comparable to the best method of binary phase shift keying, which has the highest noise immunity among existing methods. High noise immunity of the optimal method is conditioned by the fact that the transfer of information bits are used eigenvectors substrip matrices, which is known to be orthogonal to each other.

Security information transfer is provided by the permutations of the eigenvectors to the formation of the channel signal, which requires knowledge of the exact location of rearranged eigenvectors when restoring data at the receiving end, the key of this method of protection will be a map of the exact location of the eigenvectors.

The method allows to significantly improve the efficiency of frequency resources by minimizing the energy fraction of the chapels of a given frequency range, while also significantly reduce the interference between adjacent channels. In addition, the generated channel signal has a noise immunity comparable with the most noise-stable binary phase shift keying, with no loss in speed of information transmission.



## РАСШИРЕНИЕ ЧАСТОТНОГО ДИАПАЗОНА АУДИОСИГНАЛОВ НА ОСНОВЕ МОДЕЛИ ВЗВЕШЕННЫХ КЛАСТЕРОВ

Любимов Н.А., Лукин А.С.

МГУ им. М.В. Ломоносова, ф-т вычислительной математики и кибернетики,  
лаборатория математических методов обработки изображений

### Введение

*Расширение частотного диапазона (РЧД)* – это обработка аудиосигнала, направленная на улучшение качества звучания за счёт генерации отсутствующих частотных компонент сигнала [1]. Необходимость в такой обработке возникает, если аудио сигнал подвергся нежелательной фильтрации в процессе звукозаписи, передачи или воспроизведения. Методы РЧД условно можно разделить на два типа по наличию дополнительной информации об отсутствующих частотных компонентах:

- «слепые» («blind») – не использующие дополнительной информации о ресинтезируемых компонентах
- «не слепые» («non-blind») – некоторая информация о ресинтезируемых частотных компонентах передается в аудиопотоке в закодированном виде.

Широко известным примером «не слепого» метода РЧД является технология *репликации спектральной полосы* – SBR [2]. Эта технология является основой таких форматов аудио кодирования, как mp3PRO и AAC+. В них сжатию подвергается не весь частотный диапазон аудио сигнала, а только область низких и средних частот (НЧ). В области высоких частот (ВЧ) вычисляются параметры спектральной огибающей и передаются в потоке в виде небольшого объема дополнительной информации. На стороне декодера ВЧ-диапазон регенерируется путем транспозиции существующих спектральных компонент в область высоких частот. Форма спектральной огибающей корректируется за счет дополнительной информации. В силу слабой чувствительности слуха к искажениям в области ВЧ, данная технология позволяет увеличить эффективность кодирования аудио на 50%, что делает её привлекательной для многих систем широкоформатного вещания, таких как XM Satellite Radio и Digital Radio Mondiale [2].

«Слепые» методы не используют дополнительной информации, а предсказывают форму спектральной огибающей в области ВЧ на основе информации, представленной в НЧ-диапазоне. Один из простых видов такого предсказания описан в работе [3], где предлагается линейная экстраполяция спада энергии в области высоких частот. В статье [4] предлагается эффективный вычислительный метод регенерации ВЧ-диапазона на основе нелинейного искажения входного сигнала. Интересный подход предложен в [5], где экстраполяция ВЧ-огибающей осуществляется за счет линейного предсказания одновременно в частотной и временной области. Другие варианты, использующие статистические модели предсказания огибающей представлены в работах [6,7].

Общая схема работы «слепых» и «не слепых» алгоритмов выглядит одинаково. Вначале дискретный сигнал переводится в частотную область с использованием частотно-временных преобразований, таких как

онокное преобразование Фурье, квадратурные зеркальные фильтры или модифицированное дискретное косинусное преобразование. Дальнейшая обработка осуществляется в два этапа:

1. грубая генерация ВЧ-сигнала (транспозиция, нелинейное искажение),
2. формирование амплитудной огибающей спектра высоких частот.

Получившийся таким образом спектр переводится обратно во временную область. Процесс РЧД для случая генерации спектра высоких частот проиллюстрирован на Рис. 1.

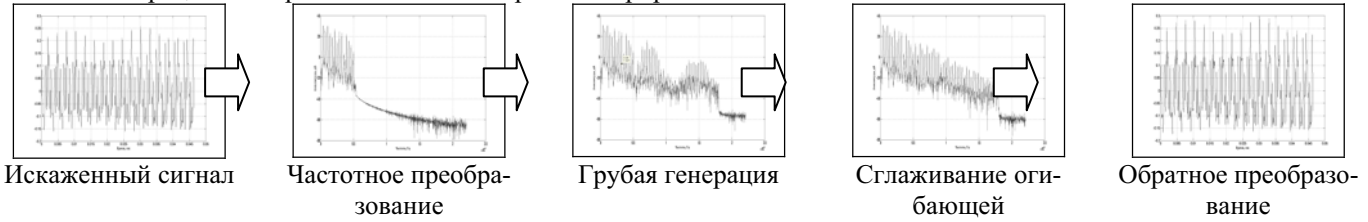


Рис. 1. Общая схема работы алгоритма РЧД.

**Предлагаемый метод**

В данной работе предложен новый метод расширения частотного диапазона (РЧД) аудиосигнала, общая схема которого аналогична описанной выше. На вход подается дискретный звуковой сигнал с ограниченным спектром частот, т.е. в спектре данного сигнала отсутствуют гармоники на частотах выше некоторого значения  $f_{cut}$ . Требуется расширить спектр сигнала, добавив в него отсутствующие частотные компоненты, не теряя при этом естественность звучания.

В качестве частотно-временного преобразования (банка фильтров) используется оконное преобразование Фурье (STFT) с окном Ханна длиной порядка 30 мс и перекрытием 15 мс. Блоки сигнала  $\mathbf{x}_t^w$  преобразуются в частотную область:  $\mathbf{y}_t = \mathbf{M}\mathbf{x}_t^w, \quad t = 1, 2, \dots, T, \quad \mathbf{M} \in R^{N \times N}$  (3)

а также подвергаются нелинейному преобразованию  $\mathbf{z}_t = F(\mathbf{x}_t^w), \quad t = 1, 2, \dots, T$  (4)

где  $F$  – нелинейная функция. Обычно  $F(x) = |x|^s, \quad s \geq 1$ . Суть нелинейной фильтрации  $F$  состоит в генерации новых гармоник в исходном сигнале. На практике для устранения эффекта наложения спектров (*алиасинга*) используется процедура передискретизации сигнала до и после нелинейного преобразования:

$$\mathbf{z}_t = \downarrow \left( \mathbf{g}_2 * F(\mathbf{g}_1 * (\uparrow \mathbf{x}_t^w)) \right) \quad (5)$$

Здесь  $\mathbf{g}_1, \mathbf{g}_2$  – антиалиасинговые фильтры. Окончательно данный сигнал переводится в частотное представление:  $\tilde{\mathbf{y}}_t = \mathbf{M}\mathbf{z}_t, \quad t = 1, 2, \dots, T$  (6)

Таким образом, для каждого входного аудиофрейма мы имеем два экземпляра дискретного спектра:  $\mathbf{y}_t$  – спектр искаженного низкочастотного сигнала и  $\tilde{\mathbf{y}}_t$  – «расширенный» спектр исходного сигнала. В таком случае процедура добавления новых спектральных компонент имеет простой вид:

$$s_t(k) = \begin{cases} y_t(k), & 1 \leq k \leq k_{cut}, \\ \tilde{y}_t(k), & k_{cut} < k \leq \frac{N}{2} + 1 \end{cases} \quad t = 1, 2, \dots, T \quad (7), \quad \text{где } k_{cut} = \left\lfloor \frac{f_{cut} N}{f_s} \right\rfloor, \text{ а } f_s - \text{ частота}$$

дискретизации сигнала. Преобразуя сигнал  $s_t$  обратно во временную область и складывая последовательно все фреймы  $t = 1, 2, \dots, T$ , мы получаем новый сигнал с расширенным частотным диапазоном.

Для улучшения качества предлагается сглаживать спектральную огибающую сгенерированного сигнала перед тем как суммировать его с исходным:  $s_t(k) = \begin{cases} y_t(k), & 1 \leq k \leq k_{cut}, \\ \rho_t(k)\tilde{y}_t(k), & k_{cut} < k \leq \frac{N}{2} + 1 \end{cases} \quad t = 1, 2, \dots, T \quad (8)$

Веса  $\rho_t(k)$  необходимо задавать такими, чтобы уровень вносимых искажений был минимальным. Очевидно, что оптимальным вариантом является выбор  $\rho_t(k) = |y_t^0(k)| / |\tilde{y}_t(k)|$ , где  $y_t^0(k)$  – исходный (неискаженный) спектр.

В данной работе предложен подход, позволяющий приблизительно оценить огибающую высоких частот на основе информации, представленной в низкочастотном диапазоне. Сигнал разбивается на 24 частотных диапазона в соответствии с критическими полосами слуха. Значения  $\rho_t(k)$  предполагаются постоянными в

пределах одной критической полосы. В каждом частотном диапазоне вычисляется логарифм мощности сигнала. Далее строится регрессионная модель, описывающая взаимосвязь характеристик спектральных огибающих. Эта взаимосвязь моделируется за счет использования *кластеров пар*, каждый из которых представляет собой линейное отображение  $q_i(x)$  пространства НЧ огибающих в ВЧ. Входные значения НЧ огибающих используются для формирования весов  $\beta_i(x)$ , связывающих различные кластеры пар. Это позволяет плавно интерполировать формы высокочастотной огибающей при переходе от одного кластера пар к другому, что является преимуществом по сравнению с моделью кластеров пар с жестко зафиксированными границами [6]. В литературе описанная регрессионная модель носит название *модель взвешенных кластеров* – CWM [8]. На рис. 2 представлен процесс генерации высокочастотных огибающих для случая 7 кластеров.

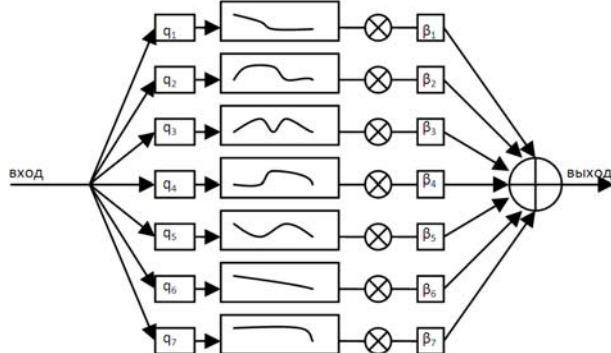


Рис. 2. Генерация высокочастотной огибающей с использованием модели взвешенных кластеров.

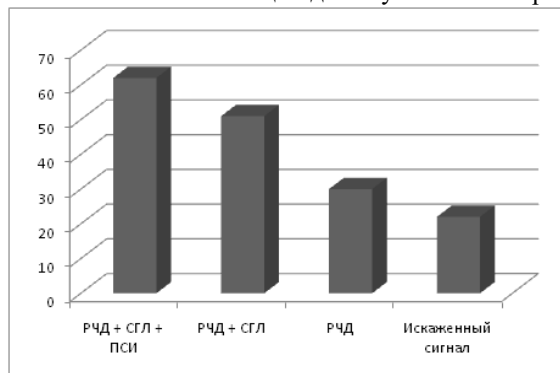


Рис. 3. Средние субъективные оценки качества различных алгоритмов РЧД.

### Результаты

Для сравнительного анализа результатов работы различных алгоритмов был проведен тест, в котором участвовали независимые слушатели. Каждому из них предлагалось сравнить фонограмму исходного (неискаженного) сигнала с фонограммой искаженного сигнала и вариантами обработки различными алгоритмами РЧД. Субъективная оценка качества ранжировалась от 0 (сильно искаженное звучание по сравнению с оригиналом, большой уровень помех) до 100 (точная копия оригинальной фонограммы, отсутствие помех). В тестовом наборе участвовали записи как речевых, так и музыкальных сигналов с различной степенью полифонии. Частота дискретизации входных данных составляла 44.1 кГц. Для моделирования высокочастотного искажения к сигналам был применен низкочастотный фильтр с частотой среза 5.5 кГц. Алгоритмы РЧД включали в себя:

- **РЧД** – регенерация ВЧ диапазона проводилась без использования сглаживания весовым коэффициентом, согласно формуле (7), аналогично методу [4].
- **РЧД + СГЛ** – регенерация ВЧ диапазона проводилась с использованием весового коэффициента для сглаживания (8). В качестве параметров спектральной огибающей бралась средняя мощность сигнала в узкой полосе частот без использования моделей восприятия.
- **РЧД + СГЛ + ПСИ** – регенерация ВЧ диапазона проводилась с использованием весового коэффициента для сглаживания (8) с учетом описанной в данной статье модели восприятия звукового сигнала.

На рис. 3 изображены средние субъективные оценки качества приведенных алгоритмов. Основным недостатком построенного алгоритма является падение качества при обработке полифонического сигнала. Предположительно, это связано с возникновением эффекта интермодуляции при нелинейном искажении сигнала. В дальнейшем планируется адаптировать методы РЧД, основанные на нелинейном искажении, к полифоническим сигналам путем предварительного разделения источников сигналов в полифонической смеси и применения описанного алгоритма РЧД отдельно к каждому источнику.

### Заключение

В данной работе был рассмотрен новый метод расширения частотного диапазона аудиосигнала в случае потери высоких частот. Для генерации высокочастотных компонент применялся метод нелинейной фильтрации. Для предсказания высокочастотной огибающей была применена модель взвешенных кластеров, которая имеет ряд преимуществ по сравнению с моделями векторного квантования. Данный метод позволил повысить субъективную оценку качества по сравнению с другими методами РЧД.

Работа выполнена при поддержке ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009–2013 годы.

### Литература

1. Erik Larsen, Ronald M. Aarts, “Audio Bandwidth Extension: Application of Psychoacoustics, Signal Processing and Loudspeaker Design”, John Wiley & Sons, ISBN: 0-470-85864-8, September 2004.

2. M. Dietz, L. Liljeryd, K. Kjorling and O. Kunz, "Spectral Band Replication, a novel approach in audio coding," 112<sup>th</sup> AES convention, Munich, Germany, May 2002.
3. Chi-Min Liu, Wen-Chieh Lee, and Han-Wen Hsu "High Frequency Reconstruction for Band-Limited Audio Signals", DAFX'03, London, UK, September 2003.
4. Manish Arora, Joonhyun Lee, and Sangil Park "High Quality Blind Bandwidth Extension of Audio for Portable Player Applications", 120<sup>th</sup> AES convention, Paris, France, May 2006.
5. Chatree Budsabathon, Akinori Nishihara "Bandwidth Extension with Hybrid Signal Extrapolation for Audio Coding", IEICE Trans. Fundamentals, Vol. E90-A, No. 8, August 2007.
6. N. Enbom and W.B.Klein "Bandwidth Expansion of Speech Based on Vector Quantization of the Mel Frequency Cepstral Coefficients", in IEEE Workshop on Speech Coding, Porvoo, Finland, 1999.
7. M. Nilsson, H. Gustafsson, S.V. Andersen, and W.B. Kleijn. "Gaussian mixture model based mutual information estimation between frequency bands in speech." ICASSP '02, IEEE International Conference on, 2002.
8. Neil Gershenfeld, "Cluster-Weighted Modeling: Probabilistic Time Series Prediction, Characterization and Synthesis", Nature of Mathematical Modeling, MIT Press, 1998.

## BANDWIDTH EXTENSION FOR AUDIO SIGNALS USING CLUSER-WEIGHTED MODELING

Lyubimov N., Lukin A.

Moscow Lomonosov State University, Dept. of Computational Mathematics and Cybernetics  
Laboratory of Mathematical Methods of Image Processing

### Introduction

*Bandwidth extension* is the process of re-synthesizing missing frequency components in order to improve the subjective quality of the audio signal. Bandwidth extension methods can be found in modern perceptual audio coding standards, such as mp3PRO and AAC+. Such methods can be blind, when no information about missing signal components is available, and non-blind, when certain information about missing components is available during the synthesis stage.

A typical algorithm flowchart for both blind and non-blind methods looks as follows:

1. Time-frequency decomposition,
2. "Rough" generation of high-frequency spectral content,
3. Shaping of the energy spectrum envelope of high-frequency content,
4. Synthesis of the resulting signal from a time-frequency representation.

### Algorithm description

In this work, a new algorithm for blind bandwidth extension is proposed. It is capable of accurate prediction of high-frequency energy envelopes using a Cluster-Weighted Model for MFCC coefficients of the audio signal.

A bandwidth-reduced audio signal  $x(t)$  is input to the algorithm. It is transformed using STFT with 30-ms Hann windows that overlap by 15 ms. A nonlinear distortion (waveshaping) is used to generate "rough" high-frequency components in time domain:  $z(t) = |x(t)|^s$ . Aliasing is reduced by the use of oversampling. These high-frequency components are transformed using a similar STFT filter bank.

To finally shape the resulting high-frequency signal, its energy is computed in 24 critical bands, and a regression model is developed to predict the shape of high-frequency energy envelope from a low-frequency energy envelope. In this work, a Cluster-Weighted Modeling is proposed for such prediction. The model is trained on the original full-bandwidth audio signals to determine a set of clusters of low-frequency envelopes and corresponding high-frequency envelopes. During evaluation stage, each input low-frequency envelope is represented as a weighted sum of several "cluster" envelopes, and the corresponding high-frequency envelope is predicted as a weighted sum of modeled high-frequency envelopes.

Once the desired high-frequency envelope is calculated, the shape of a "rough" high-frequency signal is transformed to match the desired envelope using STFT filter bank.

### Results and conclusion

The algorithm has been evaluated on speech and music using a subjective evaluation protocol with several listeners. The proposed method has been compared with other methods, such as linear extrapolation of high-frequency envelope (see full references within the paper). The proposed algorithm has shown the highest subjective quality results.

Evaluations have shown that the algorithm is more effective on speech and solo music than on polyphonic music. A possible cause of this effect is introduction of intermodulation components by the process of nonlinear distortion. In our future work, we are planning to apply source separation techniques for individual processing of signal components.