

## “КЛЕТОЧНАЯ” МОДЕЛЬ ПРЕДСТАВЛЕНИЯ АУДИАЛЬНОЙ ИНФОРМАЦИИ

Асланов А.Е.

Донецкий национальный университет, физический факультет

E-mail: aslanov@dongu.donetsk.ua

### Abstract

*Aslanov A.E. A “place” model for representation of the sound information. In given article the new approach to the analysis of the sound information is offered. This approach is based on the theory of a place offered by Hungarian Nobel winner G. von Bekesy. In the article the biological structure of an ear cochlea, the physical processes occurring in it, physiological perception of oscillatory processes and psychological interpretation by the human of the sound information are considered. The basic innovation is transition from the spectrum interpretation received by means of Fourier transformation, to interpretation of basilar membranes deflections.*

### Введение.

Известны производственные процессы, например, переработка зерна [1], тепловые электростанции [2] и другие, где принятие решения об управлении основывается на интерпретации аудиальной информации о текущем состоянии объекта управления, воспринимаемой человеком-экспертом. В развивающихся сейчас электронных технологиях и робототехнике актуальным остается проблема распознавания человеческой речи и выделение из нее смысловой нагрузки для передачи команд электронике или роботам [3]. В музыкальной и цифровой индустрии возникают проблемы повышения качества компрессии звуковых файлов и синтезировании человеческой речи, для возможного общения компьютеров с человеком.

Решение указанных проблем возможно на основе моделей восприятия аудиальной информации. Несмотря на достижения когнитивных наук в этой области, формальные модели восприятия человеком аудиальной информации отсутствуют. В данной статье предпринята попытка создания новой модели представления аудиальной информации, учитывающей физические процессы, происходящие в человеческом ухе, биологическое строение внутреннего уха, физиологические ощущения и психологическую интерпретацию звука человеком.

### Параметры аудиального сигнала.

Как известно, акустический сигнал, как и любой другой колебательный процесс, принято характеризовать такими физическими параметрами как частота, амплитуда, фаза и сложность (для сложного, многотонального сигнала). Однако, как показывают исследования по когнитивной психологии, в частности в области ощущения и восприятия [4, 5, 6], психологические параметры акустического сигнала отличаются от физических. Эта особенность связана с тем, что в психологических экспериментах по аудиальному восприятию эффект восприятия связывают не с перечисленными физическими параметрами сигнала, а с некоторыми их производными величинами.

Таковыми психологическими параметрами являются высота тона, громкость и тембр, которые можно сопоставить с частотой, амплитудой и сложностью (фаза не воспринимается человеком при моноауральном анализе).

**Теория места.**

Основой теории места является представление о том, что разные частоты возбуждают разные участки базилярной мембраны, как показано на рис.1, а как следствие, возбуждаются разные волосковые клетки, связанные с этими участками. С конкретными волосковыми клетками связаны определённые волокна слухового нерва и структуры аудиальной коры головного мозга. Колебания, которые передаются базилярной мембране, создают определенный паттерн вибраций, соответствующий уникальному виду бегущей волны в жидкости. В какой именно точке огибающая бегущей волны достигает своего пика, зависит от частоты звука. Рис.1 заимствован из [6].

Физическое подтверждение теории сделано в работах [7, 8, 9].

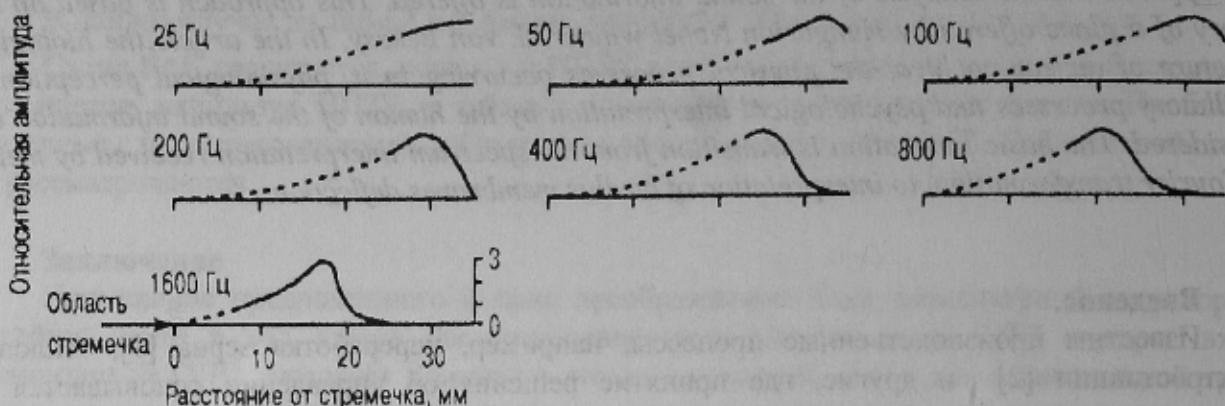


Рис.1. Зависимость формы огибающей от частоты возбуждения

**Выбор формальной модели.**

Выбор формальной модели базируется на когнитивных моделях восприятия человеком частотной структуры сигнала. В литературе [4 – 10] рассматриваются следующие модели: гармоническая октава (музыкальные ноты), мелодическая октава (шкала мелов) и частотные группы (шкала барков).

Построение формальных моделей на основе первых двух представленных выше моделей, используемых психологами, затруднительно, так как первая модель изобретена, в основном, для облегчения настройки музыкальных инструментов [4, 6, 11], а вторая построена только на сравнительном анализе одиночных тонов при помощи человека и, соответственно, от него не отделима [4, 5, 6]. Третья же, в свою очередь, представляет наиболее близкую модель к физиологическим ощущениям человека и, в какой-то степени, не зависима от него. Она может быть использована не только для анализа одиночных тонов, но и для более сложных сигналов.

Поскольку в статье рассматриваются именно сложные сигналы (технологические шумы), то для их исследования более логичным будет выбор модели частотных групп (шкалы барков).

Модель барков предложена в [4 – 10] для исследования механизмов восприятия частотной структуры сложного сигнала, образованного несколькими гармоническими составляющими. Сущность модели состоит в том, что она связывает между собой такие параметры восприятия высоты как: барк, линейные размеры базилярной мембраны (мм, мкм), степени различия высоты в безразмерных единицах, мелы и количество волосковых клеток. Эта зависимость приведена в табл.1, позаимствованной из [6].

Из табл.1 и экспериментальных данных, приведенных в литературе по когнитивной психологии, следует, что на одну степень различия по высоте приходится примерно 4 волосковые клетки базилярной мембраны. На один барк приходится примерно 35 ступеней различия и примерно 146 волосковых клеток. То есть, можно сделать предположение, что

пока аудиальное возбуждение не перейдет на следующие 4 клетки, человек не будет различать следующую ступень различия по высоте. Это одно из требований, которому должна удовлетворять модель представления высоты аудиального сигнала. В данной статье предлагается считать, что в одном барке укладывается 144 волосковые клетки, что, при группировке по 4 клетки, дает 36 ступеней различия по высоте в каждом барке, а всего клеток, с учетом 24 барков, получается 3456, что так же удовлетворяет результатам исследования восприятия аудиального сигнала, изложенными в когнитивной модели [4, 5, 6].

Таблица 1 –

| Число частотных групп | Длина отрезка кортиевого органа | Число ступеней | Тональное число мел | Число волосковых клеток |
|-----------------------|---------------------------------|----------------|---------------------|-------------------------|
| 1                     | 1,3 мм                          | 35             | 90                  | 146                     |
| 0,8                   | 1 мм                            | 27             | 70                  | 110                     |
| 0,02                  | 37 мкм                          | 1              | 2,5                 | 4,2                     |
| 0,011                 | 15 мкм                          | 0,4            | 1                   | 1,7                     |
| 0,007                 | 9 мкм                           | 0,24           | 0,6                 | 1                       |

#### Расчет стационарных границ барков.

Из анализа, приведенного в [4 – 10], был сделан вывод, что, весь частотный диапазон воспринимается человеком неоднозначно: до 500 Гц – примерно линейно, а после 500 Гц – примерно логарифмически. Также из [6, 8, 10] следует, что до 500 Гц ширина частотных групп (барков) составляет  $\approx 80$  Гц, а после 500 Гц –  $\approx 17-20\%$  относительно центральной частоты барка. Таким образом, весь частотный диапазон делят на 24 барка [6, 8, 10].

Исходя из того, что в диапазоне частот от 20 до 500 Гц укладываются шесть первых барков (1-й, ..., 6-й) по 80 Гц шириной, а в диапазоне от 500 до 20000 Гц – остальные 18 барков (7-й, ..., 24-й) с одинаковой относительной шириной – рассчитаем границы этих барков.

Границы первых 6 барков очевидны и приведены в табл.2.

Таблица 2 –

Границы первых 6-и барков в Гц

| 1-й    | 2-й     | 3-й     | 4-й     | 5-й     | 6-й     |
|--------|---------|---------|---------|---------|---------|
| 20–100 | 100–180 | 180–260 | 260–340 | 340–420 | 420–500 |

Для расчёта границ оставшихся 18 барков (с 7-го по 24-й) необходимо сделать предположение, что 20000 Гц превышает 500 Гц в 18 раз по какому-то основанию логарифма и, соответственно, рассчитать это основание по формуле:

$$\log_x \left( \frac{20000}{500} \right) = 18, \quad x = \sqrt[18]{40} \approx 1.227, \quad (1)$$

Затем остается сделать предположение, что верхние границы рассчитываемых барков превышают 500 Гц в 1, 2, ..., 18 раз и, соответственно, найти численное значение этих границ по формуле:

$$\log_{1.227} \left( \frac{y_k}{500} \right) = k, \quad y_k = 1.227^k \cdot 500, \quad (2)$$

где  $x$  – основание логарифма,  $y_k$  – верхняя граница  $k$ -го барка,  $k$  – целое число в диапазоне от 1 до 18.

Рассчитанные, округленные до целого, границы оставшихся 18 барков сведены в табл.3.

Таблица 3 –

Округленные значения границ барков с 7-го по 24-й в Гц.

|             |             |              |               |               |               |
|-------------|-------------|--------------|---------------|---------------|---------------|
| 7-й         | 8-й         | 9-й          | 10-й          | 11-й          | 12-й          |
| 500 – 614   | 614 – 753   | 753 – 925    | 925 – 1135    | 1135 – 1393   | 1393 – 1710   |
| 13-й        | 14-й        | 15-й         | 16-й          | 17-й          | 18-й          |
| 1710 – 2099 | 2099 – 2576 | 2576 – 3162  | 3162 – 3882   | 3882 – 4764   | 4764 – 5848   |
| 19-й        | 20-й        | 21-й         | 22-й          | 23-й          | 24-й          |
| 5848 – 7178 | 7178 – 8811 | 8811 – 10815 | 10815 – 13275 | 13275 – 16294 | 16294 – 20000 |

Из таблиц 2 и 3 видно, что преобразование шкалы частот в шкалу барков несет в себе огрубление информации, содержащейся в исходном сигнале, за счет широких диапазонов барков. Для решения некоторых классов задач такое огрубление не всегда подходит, например, для различения нескольких близких по частоте сигналов. Или наоборот, даже такая огрубленная модель все еще несет в себе избыточную информацию, например, для объединения различных сигналов в классы «высокий», «средний» и «низкий» звуки. Плюс ко всему, из-за того, что рассчитанные границы барков стационарны, дальнейший анализ аудиального сигнала может повлечь за собой возникновение некоторых ошибок, приводящих к расхождениям в интерпретации исходного сигнала автоматической системой и человеком-экспертом. Например, при расчете громкости сложного узкополосного сигнала, когда ширина полосы спектральных составляющих не превышает ширины одного барка, но граничные частоты этой полосы располагаются на частотной шкале так, что стационарная граничная частота двух соседних барков попадает в промежуток между граничными частотами полосы рис.2, то автоматическая система даст завышенную оценку громкости, нежели человек-эксперт. Поэтому, в данной статье для анализа аудиальных данных предлагается производить переход от обычной частотной шкалы не к шкале барков, а к более детальной шкале волосковых клеток.

На рис.2 обозначено:  $f_{n-1}, f_n, f_{n+1}$  – стационарные граничные частоты соседних барков;  $\Delta f$  – ширина полосы спектральных составляющих сложного узкополосного сигнала;  $L$ , дБ – уровень амплитуды в децибелах;  $f$ , Гц – частота в герцах.

Если принять во внимание, что  $\Delta f \leq (f_{n+1} - f_n)$  или  $\Delta f \leq (f_n - f_{n-1})$ , то при уровне амплитуды в 40 дБ (предположим, что  $f_n = 1000$  Гц) громкость такого сигнала человеком-экспертом будет оценена как 1 сон, а автоматическая система может оценить громкость такого сигнала в 2 сон, из-за того, что захватываются два барка по 40 дБ в каждом.

### Расчет стационарных границ волосковых клеток.

Для дальнейшей работы нам необходимо рассчитать частотные диапазоны, в которых будут моделироваться волосковые клетки.

Принимая во внимание тот факт, что в модели присутствует 3456 клеток, в один барк входит 144 клетки, а до 500 Гц укладывается 6 барков, получим, что первые  $6 \times 144 = 864$  волосковые клетки будут моделироваться с фиксированной шириной диапазона частот  $(500 - 20) / 864 \approx 0.56$  Гц. Остальные  $3456 - 864 = 2592$  волосковые клетки моделируются с

одинакової відносної ширини діапазона частот, розраховуваної по алгоритму, використовуваним при розрахунку формул (1) і (2).

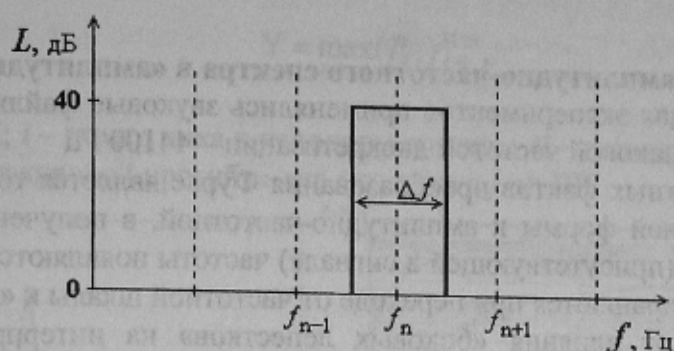


Рис.2. Спектр узкополосного сигнала и стационарные границы барков

$$\log_{x_{\text{вк}}} \left( \frac{20000}{500} \right) = 2592, \quad x_{\text{вк}} = \sqrt[2592]{40} \approx 1.0014, \quad (3)$$

$$\log_{1.0014} \left( \frac{y_{\text{вк}l}}{500} \right) = 1, \quad y_{\text{вк}l} = 1.0014^l \cdot 500 \quad (4)$$

где  $x_{\text{вк}}$  – новое основание логарифма для волосковых клеток,  $y_{\text{вк}l}$  – верхняя граница  $l$ -ых волосковых клеток,  $l$  – целое число в диапазоне от 1 до 2592.

В результате получается множество

$$Y = \{y_{\text{вк}j}\}_{j=1}^{3457}, \quad (5)$$

содержащее границы частотных диапазонов, в которых реагируют волосковые клетки.

Причем,

$$y_{\text{вк}j+1} - y_{\text{вк}j} \approx 0,56 \text{ Гц} \quad \text{при} \quad 1 \leq j \leq 865$$

$$\log_{x_{\text{вк}}} \left( \frac{y_{\text{вк}j+1}}{y_{\text{вк}j}} \right) = 1 \quad \text{при} \quad 865 \leq j \leq 3457 \quad (6)$$

где  $x_{\text{вк}}$  – то же, что и в (3);  $y_{\text{вк}1} = 20 \text{ Гц}$ , а  $y_{\text{вк}3457} = 20000 \text{ Гц}$ .

### Выбор параметров для преобразования Фурье.

Из (6) видно, что минимальная ширина диапазона частот, в которой реагируют волосковые клетки (ВК), равна 0,56 Гц. Таким образом, чтобы получившийся спектр имел дискретность, достаточную для фиксирования значений на минимальных диапазонах ВК, необходимо специальным образом подобрать окно анализа для преобразования Фурье. Кроме того, желательно, чтобы в этот диапазон попадали не менее 2-х частотных составляющих, для возможности расчета амплитуды во всем диапазоне. Авторами было предложено использовать дискретность в частотном диапазоне равную 0,1 Гц, что дает наличие  $\geq 5$ -и элементов в каждом диапазоне реагирования ВК.

Однако выбор дискретности в 0,1 Гц влечет за собой необходимость выбора окна анализа для преобразования Фурье длиной в 10 с. Такое не всегда можно сделать, так как длительность квазистационарных участков реальных сигналов зачастую намного меньше.

Исходя из этого, автором было предложено воспользоваться способом, описанным в [12] для уменьшения шага дискретности по частоте при неизменной длине окна анализа реального сигнала.

### **Преобразование амплитудно-частотного спектра в «амплитудно-клеточный».**

В нашей работе для экспериментов применялись звуковые файлы разной структуры и длительности, но с одинаковой частотой дискретизации – 44100 Гц

Одним из неприятных фактов преобразования Фурье является то, что, после перехода от амплитудно-временной формы к амплитудно-частотной, в полученном спектре с обеих сторон от резонансной (присутствующей в сигнале) частоты появляются «боковые лепестки» [12, 13, 14], которые сохраняются при переходе от частотной шкалы к «клеточной».

Для избавления от влияния «боковых лепестков» на интерпретацию аудиального сигнала авторами было предложено сглаживать кривую спектра, полученную после преобразования Фурье. Сглаживание происходит путем «соединения» пиков локальных максимумов («боковых лепестков») и интерполяции значений между ними рис.3.

На рис.3.а сплошной линией обозначена кривая, полученная после преобразования Фурье, а пунктиром – кривая после сглаживания.

После преобразования Фурье исходного сигнала с частотой дискретизации 44100 Гц, окном анализа 0,1 с., вставленного в окно 10 с., получается набор из 220500 значений амплитуды для частотного диапазона от 0,1 Гц до 22050 Гц с шагом дискретности 0,1 Гц. Так как полученный диапазон получился избыточным, то из него мы выделяем только элементы из слышимого диапазона, а это элементы с номерами от 200 до 200000. Далее этот диапазон разбивается на 3456 частотных полос, соответствующих 3456 ВК. Из каждой такой полосы выбирается максимальное значение амплитуды и ставится в соответствие той клетке, которая соответствует данной полосе. В результате мы получаем «амплитудно-клеточный» спектр.

### **«Кривые прогиба» базилярной мембраны.**

В данной работе мы сделали предположение, что каждый пик, полученный после преобразования спектра в «клеточную» шкалу, «прогибает» базилярную мембрану в зависимости от номера клетки и своей амплитуды. То есть, форма «кривой прогиба» (на рис.3.б шкала амплитуд приведена в логарифмическом масштабе для лучшего отображения кривых) зависит от «слышимой» амплитуды, а «слышимая» амплитуда зависит от реальной амплитуды и номера клетки. А так же от кривых одинаковой громкости, преобразованных для «клеточной» шкалы.

Прежде чем перейти к описанию алгоритма расчета «кривых прогибов», необходимо сказать, что форма прогиба базилярной мембраны напрямую зависит от амплитуды частотной составляющей, входящей в сигнал [8, 9]. В этом источнике говорится, что, при уровне громкости однотонового сигнала в 120 фон, ширина диапазона чувствительности не превосходит 3-х барков, а при уровне громкости в 0 фон – можно сделать предположение, что в реакции участвует всего 1-2 ВК. Однако, такое утверждение верно только для простых однотоновых сигналов. Мы же, в свою очередь, попытаемся предложить обобщенный метод и для многотоновых сигналов.

Что касается однотоновых сигналов, то здесь все просто. 1) После получения «клеточного» спектра мы выявляем, в какой клетке находится пик. 2) Смотрим значение реальной полученной амплитуды в этой клетке. 3) Смотрим, какому «слышимому» значению соответствует реальная амплитуда на данной клетке. 4) В зависимости от значения «слышимой» амплитуды строится кривая прогиба с максимумом в исследуемой клетке.

Что же касается многотоновых сигналов, то мы делаем предположение, что, каждый пик частотных составляющих прогибает базилярную мембрану, так же, как если бы он был один в сигнале. Поэтому, для каждого выделенного пика мы повторяем шаги 1), ... , 4) из

предыдущего алгоритма, а в конце производим операцию максимизации по всем полученным кривым для каждой клетки (рис.4.а, б) по формуле (7).

$$Y = \max_{i=1..N} (P_{i,j})_{j=1}^{3456}, \quad (7)$$

где  $j$  – номер ВК;  $i$  – номер пика в исходном спектре;  $N$  – количество пиков в исходном спектре;  $P_{ij}$  – значение «кривой прогиба» для  $i$ -го пика в  $j$ -ой ВК.

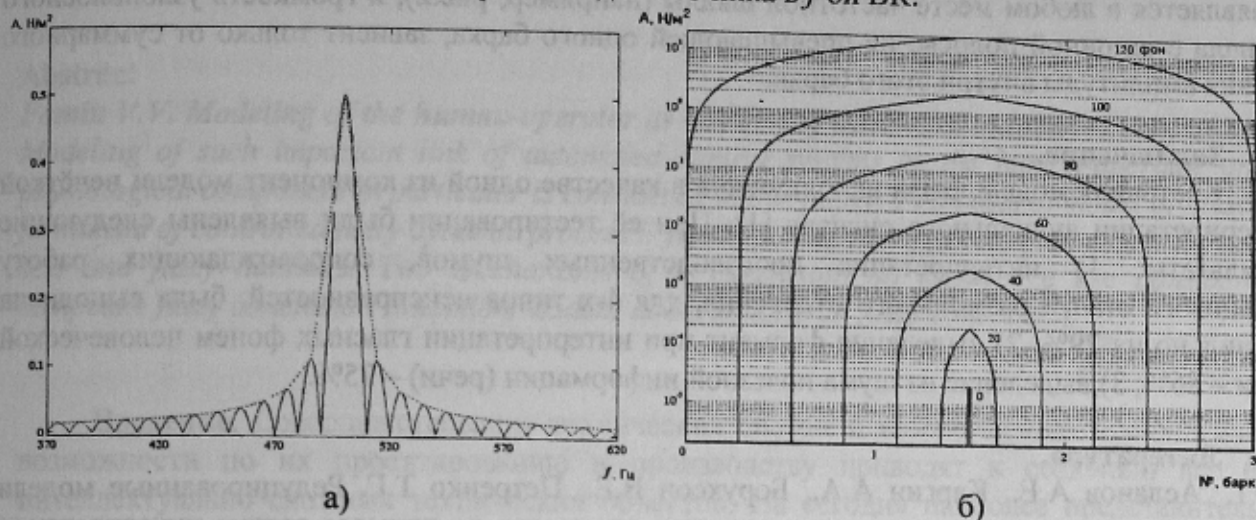


Рис.3. а) Сглаживание «боковых лепестков»; б) «Кривые прогибов» базилярной мембраны.

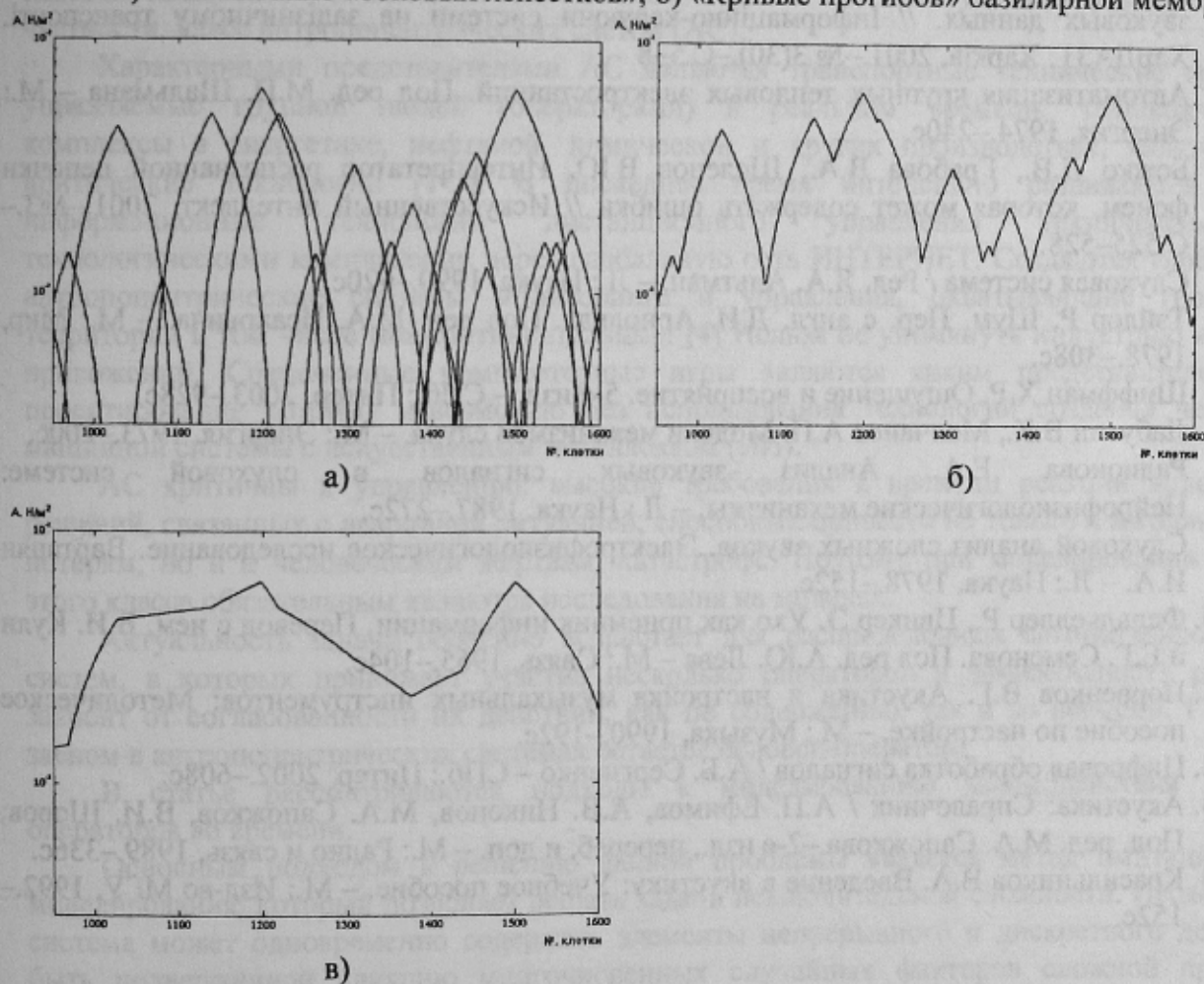


Рис.4. Фрагменты последовательного преобразования «кривых прогибов» для многотонального сигнала

Заключительным шагом в получении «клеточного» спектра является объединение пиков «кривых прогибов» (рис.4.в), расстояние между которыми меньше одного барка. Такое объединение преследует, по крайней мере, две цели.

1) Несколько близко стоящих пиков объединятся в один и, таким образом, по нашему мнению, облегчат дальнейшую интерпретацию аудиального сигнала.

2) Облегчение, на наш взгляд, дальнейшего расчета громкости, основанного на том, что понятие критической полосы (барка) не предусматривает стационарные границы и проявляется в любом месте частотной шкалы (например, рис.2), и громкость узкополосного сигнала с шириной полосы, не превышающей одного барка, зависит только от суммарного уровня амплитуды внутри этого барка.

### **Заключение.**

Клеточная модель была использована в качестве одной из компонент модели нечёткой интерпретации аудиального сигнала [1]. При её тестировании были выявлены следующие результаты: 1) интерпретация производственных шумов, сопровождающих работу вальцового станка размольного отделения, для 4-х типов неисправностей, была выполнена правильно на 79%; 2) выделение формант при интерпретации гласных фонем человеческой речи – 86%; 3) выделение из шума полезной информации (речи) – 75%.

### **Литература.**

1. Асланов А.Е., Каргин А.А., Борухсон В.Е., Петренко Т.Г. Редуцированные модели представления динамических нечетких процессов на примере задачи интерпретации звуковых данных. // Інформаційно-керуючі системи на залізничному транспорті. ХарДАЗТ, Харків, 2001.–№ 3(30).–С.5–8
2. Автоматизация крупных тепловых электростанций. Под ред. М.П. Шальмана – М.: Энергия, 1974.–240с.
3. Божко Д.В., Грабова В.А., Шелепов В.Ю. Интерпретатор распознанной цепочки фонем, которая может содержать ошибки // Искусственный интеллект, 2001.–№3.– С.525–528.
4. Слуховая система / Ред. Я.А. Альтман. – Л.: Наука, 1990.–620с.
5. Тэйлор Р. Шум. Пер. с англ. Д.И. Арнольда. Под ред. М.А. Исаковича. – М.: Мир, 1978.–308с.
6. Шиффман Х.Р. Ощущение и восприятие. 5-е изд. – СПб.: Питер, 2003.–928с.
7. Лабутин В.К., Молчанов А.П. Модели механизмов слуха. – М.: Энергия, 1973.–200с.
8. Радионова Е.А. Анализ звуковых сигналов в слуховой системе: Нейрофизиологические механизмы. – Л.: Наука. 1987.–272с.
9. Слуховой анализ сложных звуков. Электрофизиологическое исследование. Вартанян И.А. – Л.: Наука, 1978.–152с.
10. Фельдкеллер Р., Цвикер Э. Ухо как приемник информации. Перевод с нем. В.И. Кули и Е.Г. Семенова. Под ред. А.Ю. Лева – М.: Связь, 1965.–104с.
11. Порвенков В.Г. Акустика и настройка музыкальных инструментов: Методическое пособие по настройке. – М.: Музыка, 1990.–192с.
12. Цифровая обработка сигналов / А.Б. Сергиенко – СПб.: Питер, 2002.–608с.
13. Акустика: Справочник / А.П. Ефимов, А.В. Никонов, М.А. Сапожков, В.И. Шоров; Под ред. М.А. Сапожкова.–2-е изд., перераб. и доп. – М.: Радио и связь, 1989.–336с.
14. Красильников В.А. Введение в акустику: Учебное пособие. – М.: Изд-во МГУ, 1992.–152с.