

УДК 004.934

*В. Н. Поздин, М. Г. Хохлов*

## РАСЧЕТ ФОРМАНТ ПО УЧАСТКУ РЕЧЕВОГО СИГНАЛА

В статье рассматриваются проблемы, возникающие при анализе речевого сигнала. Описываются алгоритмы нахождения формант, в зависимости от используемых фильтров. Современные средства позволяют реализовать голосовое управление, голосовой ввод текста, голосовой поиск, аутентификацию и многое другое. Нахождение формант по участку речевого сигнала является одним из наиболее актуальных методов распознавания речи. Он позволяет определить содержание текста, а также аутентифицировать самого говорящего. Форманта рассматривается как амплитудный всплеск на графике спектра. При произнесении каждого звука образуется до шести формант, каждая из которых характеризует различные аспекты речи. Спектральная характеристика получается путем использования быстрого преобразования Фурье. В статье рассматриваются алгоритмы нахождения формант. Процесс выделения формант предполагает нахождение из всего множества пиков спектрального среза только тех, которые подходят по определению к формантам. Применение различных алгоритмов зависит от используемых для подавления шумов фильтров.

*Ключевые слова:* форманты, речевой сигнал, спектр, звук, быстрое преобразование Фурье, фильтр частот.

Распознавание речи с каждым годом получает все более широкое применение. Современные средства позволяют реализовать голосовое управление, голосовой ввод текста, голосовой поиск, аутентификацию и многое другое.

Нахождение формант по участку речевого сигнала является одним из наиболее актуальных методов распознавания речи. Он позволяет определить, что произнес человек, а также аутентифицировать самого говорящего.

Форманта – это амплитудный всплеск на графике спектра. При произнесении каждого звука образуется до шести формант. Первые две форманты важны

для определения произнесенного звука, остальные же несут информацию о дикторе. На рисунке 1 показан спектр русской гласной «е», отображающий 5 формант.

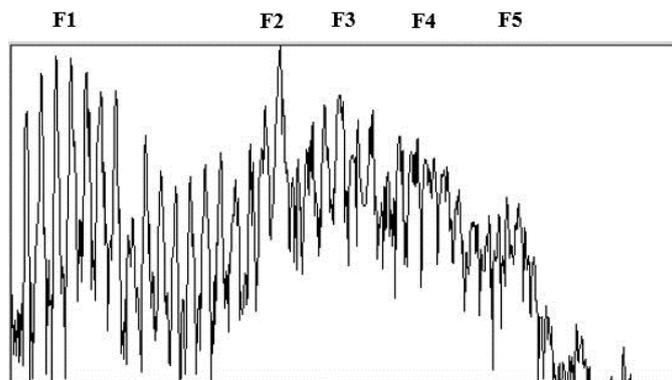


Рис. 1. Спектр гласной «е»

Спектральная характеристика получается путем использования быстрого преобразования Фурье (БПФ), результатом которого является представление сигнала в виде  $n$  частот:

$$F = \{F_1, F_2, \dots, F_n\}$$

Обозначим через  $F_{\max}$  верхнюю границу частотного спектра, а через  $F_d = 2F_{\max}$  частоту дискретизации (по теореме Котельникова). Каждому компоненту  $F_i \in F$  соответствует относительная амплитуда  $A_i$ . Компоненты  $F_i \in F$  определяются по следующей формуле:

$$F_i = \frac{iF_d}{n}, \text{ где } i = 1, 2, \dots, n \quad (1)$$

Отобразив результаты БПФ для данного момента времени, получаем графическое представление моментального спектра.

Определим пик как максимум интенсивности энергии сигнала в определенном интервале  $d$  на оси частот и выразим функцию  $P(F_k, d) = 1$  проверки максимума в интервале  $d$  следующим образом:

## Технические науки

$$P(F_k, d) = \begin{cases} 1, A_i > \max_{k-d \leq j \leq k+d} A_j, k \neq j \\ 0, A_i \leq \max_{k-d \leq j \leq k+d} A_j, k \neq j \end{cases} \quad (2)$$

Таким образом, нахождение всех пиков сводится к нахождению частот разложения  $F_i \in F$ , для которых выполняется условие  $P(F_k, d) = 1$ .

Очевидно, что процесс выделения формант предполагает нахождение из всего множества пиков спектрального среза только тех, которые подходят по определению к формантам, для чего необходимо более многостороннее рассмотрение «поведения» формант, характерных для речевого сигнала.

Одной из особенностей речевого сигнала является кратность частот формант в случае узкополосного анализатора, т.е. частота следующей форманты больше частоты предыдущей на величину, равную частоте первой форманты:

$$F_k = kF_p \quad (3)$$

Данная особенность дает возможность практически исключить нахождение ложных формант и существенно сократить количество выполняемых операций, т.к. после нахождения  $F_1$  можно лишь проверять наличие пика в точках ожидания, вычисленных по формуле (3).

Еще одна особенность речевого сигнала заключается в том, что первая форманта, в основном, лежит в диапазоне от 70 Гц до 300 Гц. Сочетание этих особенностей с учетом отношения амплитуд истинных формант к остальным пикам позволяет практически исключить ложные пик.

Таким образом, алгоритм нахождения формант:

1. Используя формулу (2) находим первую форманту.
2. Согласно формуле (3) находим точку ожидания следующей форманты.
3. Используя условие (2) для определения пика, находим следующую форманту.
4. Если не найдено необходимое количество формант, переход к шагу 2.

Данный способ основан на безошибочном нахождении первой форманты, что на практике удается не всегда. Для подавления шумов зачастую применяют поло-

совые фильтры, подавляющие частоты ниже 300 Гц и выше 4000 Гц. Первая форманта, особенно для мужского голоса, во многих случаях лежит ниже 300 Гц. В следствии чего первая форманта (либо первые две форманты) полностью теряется на спектре и найти ее амплитуду невозможно, либо происходит смещение пика.

В первом случае, как видно из рисунка 2, произойдет смещение индекса формант, т.е. третья форманта будет принята за первую.

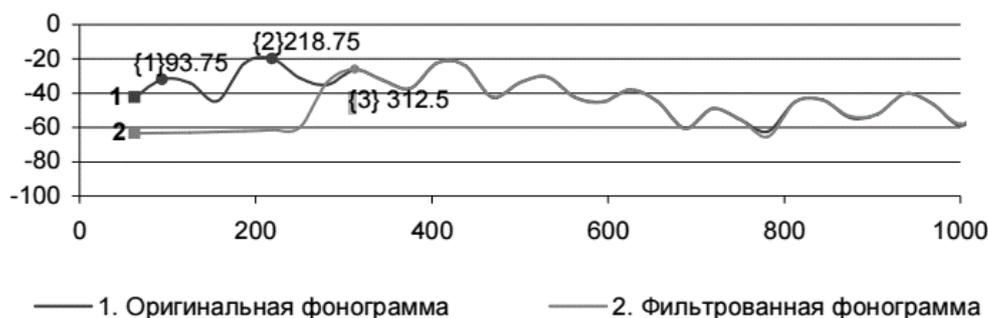


Рис. 2. Потеря двух первых пиков при фильтрации

Во втором случае (рисунок 3) от первой форманты {1} с большой интенсивностью остается смещенный по частоте "след" форманты {2}, который может быть принят за первую форманту.

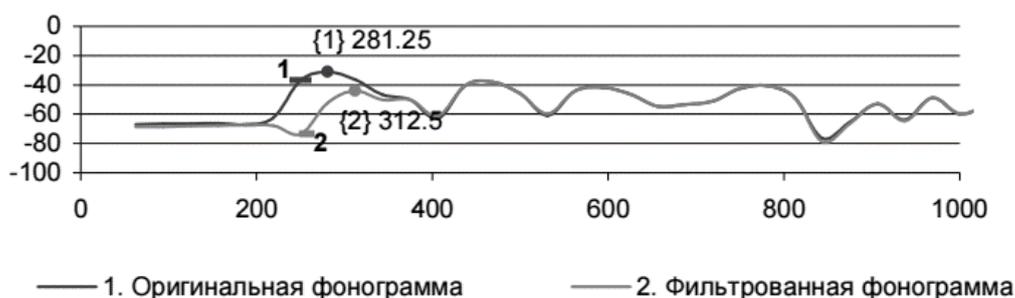


Рис. 3. Смещение пика при фильтрации

Очевидно, что потеря первой форманты (или первых двух формант) делает невозможным использование формулы (3), а сдвиг первой форманты приведет к нарушению их кратности. Следовательно, даже в случае правильного нахождения следующего пика будет получен неверный коэффициент кратности и, как следствие, будут вычислены неверные частоты ожидания следующих формант.

Во избежание данных ошибок необходимо воспользоваться другим алгоритмом.

Пусть  $F = \{F_1, F_2, \dots, F_{im}\}$  является множеством пиков моментального спектра, удовлетворяющих условию (2). Назовем расстоянием между двумя пиками количество элементарных частотных полос БПФ между ними:

$$d_{ijk} = \frac{|F_{ij} - F_{ik}|}{S}, \text{ где } S = \frac{F_d}{n}$$

Используя (1), получим:

$$d_{ijk} = \frac{\left| i_j \frac{F_d}{n} - i_k \frac{F_d}{n} \right|}{\frac{F_d}{n}} = |i_j - i_k| = i_k - i_j \text{ при } j < k .$$

Предположим, что для пары  $(i_j, i_k)$  имеют место  $i_k > i_j$

Очевидно, что при интервале проверки пика, равном  $d$ , согласно (2),  $d < d_{ijk} < n$ .

Определим граф  $G_\Delta (V, E_\Delta)$  для множества пиков  $F = \{F_{i_1}, F_{i_2}, \dots, F_{i_m}\}$  следующим образом:

$$V = \{i_1, i_2, \dots, i_m\}$$

$$E_{\Delta} = \{(i_j, i_k) \mid d_{ijk} = \Delta\}, \text{ где } \Delta = 1, 2, \dots, m$$

Для фиксированной вершины ребер  $(i_j, i_k) \in E_\Delta$  можно представить в виде:

$$i_j = \Delta (q + r), \text{ т.е. } r (i_j \bmod \Delta, \text{ где } r = 0, 1, \dots, \Delta - 1$$

Тем самым, остаток  $r = 0, 1, \dots, \Delta - 1$  разбивает множество ребер  $E_\Delta$  на непересекающиеся классы  $E_{\Delta\sigma}$ . Такие классы называются однородными.

Можно легко доказать, что однородные классы  $E_{\Delta\sigma}$  состоят из компонентов связности, являющихся простыми цепями. Используя данное утверждение, можно разработать линейный алгоритм выделения цепей в  $E_{\Delta\sigma}$ .

Назовем  $W_T = \frac{1}{|V_T|} \sum_{i_j \in T} A_{i_j}$  весом цепи  $T$ .

Назовем формантной цепью цепь с наименьшим индексом первой вершины, состоящую из более чем двух вершин и имеющую максимальный вес.

Шаги алгоритма нахождения формант:

1. Выделить все пики спектрального среза (2). Обозначить множество найденных пиков спектрального среза  $F = \{F_{i_1}, F_{i_2}, \dots, F_{i_m}\}$ .
2. Построить граф  $G_\Delta(V, E_\Delta)$  для полученного множества  $F$ .
3. Выделить цепи из множеств  $E_{\Delta\sigma}$ .
4. Выбрать формантную цепь.

Таким образом, при построении систем распознавания речи стоит уделять внимание не только алгоритмам нахождения формант, но и используемым для подавления шумов фильтрам. Первый алгоритм стоит применять в тех случаях, когда в речевом сигнале шумов нет, либо они незначительный. Второй алгоритм хорошо подходит, когда используется фильтр верхних частот с частотой среза выше 70 Гц.

**ПОЗДИН Владимир Николаевич** – кандидат технических наук, доцент кафедры автоматике и телемеханики, Вятский государственный университет. 610000, г. Киров, ул. Московская, 36.

E-mail: usr00324@vyatsu.ru

**ХОХЛОВ Михаил Геннадьевич** – магистрант, Вятский государственный университет. 610000, г. Киров, ул. Московская, 36.

E-mail: khokhlovnikha@gmail.com