ISSN 0002-306X. Изв. НАН РА и ГИУА. Сер. ТН. 2006. Т. LIX, № 1.

УДК 517.587

ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА
И ИНФОРМАТИКА

А.Г. ОГАНЕСЯН

ЭФФЕКТИВНЫЙ АЛГОРИТМ ВЫДЕЛЕНИЯ ФОРМАНТ ИЗ СПЕКТРА РЕЧЕВОГО СИГНАЛА

Разработан эффективный алгоритм выделения формант из спектра речевого сигнала в каналах тональной частоты. Используя особенности поведения формант, можно с большой точностью и оперативностью находить форманты речевого сигнала в каналах связи тональной частоты с полосой пропускания 300…3400 *Гц*. Алгоритм может быть использован в качестве инструмента в устройствах идентификации личности по голосу.

Ключевые слова: форманта, речевой сигнал, спектр, идентификация.

Введение. Одним из основных направлений в развитии речевых технологий в последние годы стала разработка методов выделения параметров речи из дискретизированных речевых сигналов как с целью передачи ее в цифровом виде в реальном масштабе времени, обеспечивающем высокую помехозащищенность и качество, так и с целью идентификации голоса. Если эту задачу применительно к передачам речевых сообщений в целом можно считать решенной аппаратно-программными средствами с применением вокодеров на базе цифровых сигнальных процессоров (ЦСП) [1], то поиск эффективных методов выделения параметров речевого сигнала сугубо программными средствами с целью оперативной идентификации и верификации голоса сохраняет актуальность [2]. При решении обеих задач в качестве основного инструмента используется спектральный анализатор, формирующий непрерывный спектр частот для каждого момента времени в координатах амплитуда – частота. Такое представление называется «*моментальным спектром*» или «*частотным срезом*» [3]. Важнейшим параметром спектра речевого сигнала является *форманта*, которую принято определять как концентрацию энергии в ограниченной частотной области. Форманта характеризуется частотой, шириной и амплитудой [4]. Такое определение форманты приводит к зависимости ее характеристик от ширины полосы анализатора, т.е. при сужении полосы количество формант может увеличиться, а при расширении - уменьшиться за счет объединения нескольких формант в одну. В реализации описываемого алгоритма использовались фонограммы с частотой дискретизации 8 *кГц* и быстрое преобразование Фурье (БПФ) размером 256, соответственно ширина полосы анализатора составляла 31,25 *Гц*.

Выделение формант сигналов, передаваемых по каналам связи тональной частоты (ТЧ), затруднено в силу специфики частотного среза этих сигналов, привносимой ограничением полосы частот. В настоящей статье описывается новый подход к решению указанной задачи применительно к сигналам с ограниченной полосой частот, отличающийся небольшими затратами времени и обеспечением приемлемой точности при выделении формант. Разработан

алгоритм, реализующий указанный подход, получена оценка сложности алгоритма и проведен эксперимент, подтверждающий правильность нахождения формант.

Подход к решению задачи. Для получения спектральной характеристики аналогового сигнала, представленного в дискретной форме, как правило, используют БПФ, результатом которого является представление сигнала в виде *n* частот

$$F = \{F_1, F_2, \ldots, F_n\}.$$

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

$$F_i = iF_d / n, \quad i = 1, 2, \ldots, n. \tag{1}$$

Отобразив результаты БПФ для данного момента времени, где ось абсцисс соответствует частоте, а ось ординат - амплитуде, можно получить графическое представление моментального спектра [3].

Огибающая линия моментального спектра, как правило, содержит большое число всплесков (пиков) отдельных частот, однако большая часть их, отличающихся небольшой амплитудой и отсутствием периодичности, неинформативна. Такие пики, как правило, находятся в области относительно высокочастотных составляющих спектра, лежащих выше 1500 *Гц*, и для речевой информации практически избыточны. Основную речевую информацию несут в себе пики с относительно большой амплитудой и очевидной периодичностью с периодом в диапазоне от 70 *Гц* до 900 *Гц*. Именно эти пики огибающей моментального спектра являются составляющими формантных линий, позволяющих производить идентификацию и верификацию голосовых сообщений. На рис 1 приведено трехмерное изображение нескольких моментальных спектров, полученных для последовательных моментов времени реальной речи. Линии, соединяющие соответственные пики моментальных спектров, образуют формантные линии.

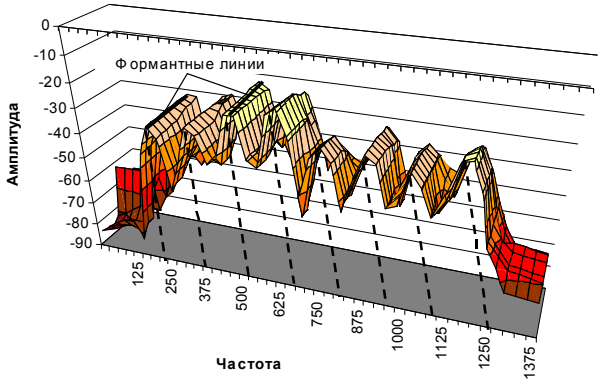

Рис 1. Образование формантных линий

Влияние шумовых источников на речевой сигнал может привести к образованию пиков моментального спектра, не являющихся частью речевого сигнала. Такие пики называются ложными.

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

$$P(F_k, d) = \begin{cases} 1, A_k > \max\limits_{k-d \le j \le k+d} A_j, \ k \neq j, \\ 0, A_k \le \max\limits_{k-d \le j \le k+d} A_j, \ k \neq j. \end{cases} \quad (2)$$

Тогда нахождение всех пиков сведется к нахождению частот разложения $F_i \in F$, для которых выполняется условие $P(F_i, d) = 1$. Назовем данный способ нахождения пиков последовательным проходом.

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» [4] формант, характерных для речевого сигнала.

Одной из особенностей речевого сигнала является кратность частот формант в случае узкополосного анализатора [3], т.е. частота следующей форманты больше частоты предыдущей на величину, равную частоте первой форманты:

$$\widehat{F}_k = k\widehat{F}_1, \qquad \widehat{F}_1 \text{ - частота первой форманты.} \qquad (3)^{\perp}$$

Данная особенность дает возможность практически исключить нахождение ложных формант и существенно сократить количество выполняемых операций, т.к. после нахождения $\widehat{F}_1$ ┌ 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* [4]. Сочетание этих особенностей с учетом отношения амплитуд истинных формант к остальным пикам позволяет практически исключить ложные пик

Для применения формулы (3) необходимо знать частоту первой форманты, для нахождения которой можно воспользоваться способом определения пика (2). После нахождения первой форманты остальные необходимо искать в точках "ожидания" согласно формуле (3), с дальнейшей проверкой условия (2). В идеале, правильно найдя частоту первой форманты, можно найти все форманты на частотном срезе. Хотя на практике кратность частот формант не всегда выполняется точно, в особенности для высоких частот, этот подход довольно эффективен и дает хорошие результаты для широкополосного сигнала, не имеющего существенного подавления частотных составляющих выше 70 *Гц*.

Попытка применения приведенного способа выделения формант к речевым сигналам, передаваемым по наиболее распространенным каналам тональной частоты (ТЧ), встречается с проблемой безошибочного нахождения первой форманты. Действительно, стандарт на каналы ТЧ, с целью подавления

шумов, вводит ограничение на полосу пропускания путем фильтрации сигнала полосовыми фильтрами, подавляющими частоты ниже 300 *Гц* и выше 3400 *Гц*. С другой стороны, во многих случаях первая форманта человеческой речи, в особенности для мужского голоса, лежит ниже 300 *Гц*, что существенно усложняет нахождение первой форманты.

Рассмотрим два случая, при которых невозможно непосредственное нахождение первой форманты сигнала, передаваемого по каналам ТЧ.

Первый случай, когда в результате фильтрации полностью теряется первый пик. На рис 2 приведены моментальные спектры одного и того же сигнала 1 - без фильтрации (сверху) и 2 - с подавлением частотных составляющих ниже 300 *Гц* (снизу).

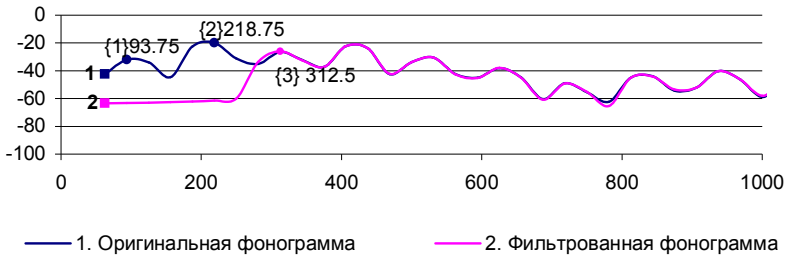

Рис 2. Потеря первого пика при фильтрации

Как видно из рис 2, первые две форманты после фильтрации полностью «исчезли», следовательно, применение способа (2) приведет к смещению индексов формант, а именно, за первую будет принята третья форманта.

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

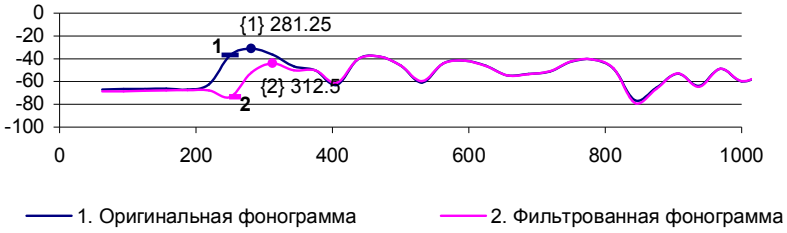

Рис 3. Смещение пика при фильтрации

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

В обоих случаях можно избежать ошибок, если коэффициент кратности определять по правильно найденным соседним пикам, лежащим выше (300+300/2) *Гц* = =450 *Гц*, где 300 *Гц* – граничная частота полосового фильтра.

Пусть $\tilde{F} = \{F_{i_1}, F_{i_2}, ..., F_{i_m}\}$ является множеством пиков моментального спектра, удовлетворяющих условию (2).

Определение 1. Назовем расстоянием между двумя пиками количество элементарных частотных полос БПФ между ними:

$d_{i_j i_k} = \frac{\left|F_{i_j} - F_{i_k}\right|}{s}$, где $s = \frac{F_d}{n}$ (величина элементарной частотной полосы БПФ).

Используя (1), получим $d_{i_j i_k} = \frac{\left|i_j \frac{F_d}{n} - i_k \frac{F_d}{n}\right|}{\frac{F_d}{n}} = \left|i_j - i_k\right| = i_k - i_j,\ при\ j < k$.

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

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

Определим граф $G_\Delta(V, E_\Delta)$[5] для множества пиков $\tilde{F} = \{F_{i_1}, F_{i_2}, ..., F_{i_m}\}$ следующим образом:

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

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

Для фиксированного (вершины ребер $(i_j, i_k) \in E_\Delta$ можно представить в виде $i_j = \Delta \cdot q + r$, т.е. $r \equiv i_j \bmod \Delta$, где $r = 0, 1, ..., \Delta - 1$.

Тем самым, остаток $r = 0, 1, ..., \Delta - 1$ разбивает множество ребер E_Δ на непересекающиеся классы $E_{\Delta(r)}$. Назовем их однородными классами.

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

Определение 2. Назовем $W_T = \frac{1}{\left|V_T\right|} \sum_{i_j \in T} A_{i_j}$ весом цепи Т.

Определение 3. Назовем *формантной цепью* цепь с наименьшим индексом первой вершины, состоящую из более чем двух вершин и имеющую максимальный вес.

Описание шагов алгоритма нахождения формант А

А1. Выделить все пики спектрального среза (2) . Обозначить множество найденных пиков спектрального среза $\tilde{F} = \{F_{i_1}, F_{i_2}, ..., F_{i_m}\}$, где $i_1 < i_2 < \cdots < i_{m-1} < i_m$ и $m << n$.

А2. Построить граф $G_\Delta(V, E_\Delta)$[5] для полученного множества $\tilde{F}$.

А3. Выделить цепи из множеств $E_{\Delta(r)}$.

А4. Выбрать *формантную цепь*.

Оценка алгоритма. Для оценки сложности алгоритма в целом оценим сложность отдельных частей алгоритма. Критерием оценки является зависимость количества элементарных операций от количества частот разбиения БПФ *n*.

Оценим сложность алгоритма по шагам:

$$\wp(A1) = O(n),$$

$$\wp(A2) = O(n^2),$$

$$\wp(A3) = O\left(\sum_{\Delta=\Delta_{min}}^{\Delta_{max}} \left|E_\Delta\right|\right) = O(m^2) \le O(n^2),$$

$$\wp(A4) = O(n^2).$$

Таким образом, сложность разработанного алгоритма не превышает

$$\wp(A) = \wp(A1) + \wp(A2) + \wp(A3) + \wp(A4) = O(n^2).$$

Для проверки эффективности алгоритма был проведен эксперимент с участием 5 специально выбранных дикторов, имеющих разные тембры голоса. В качестве тестовой фонограммы использовалась запись, содержащая всевозможные вокализованные звуки. Результаты эксперимента показали, что вне зависимости от диктора процент правильного нахождения частот первых четырех формант составил в среднем 95%.

Выводы. Как уже отмечалось, в результате прохождения по каналам тональной частоты спектр речевого сигнала терпит существенные изменения. Искажение наиболее важной низкочастотной части спектра приводит к смещению или потере первых пиков частотного среза, тем самым затрудняя использование алгоритмов выделения формант, использующих лишь принцип их кратности. Предложенный алгоритм, объединяющий способ последовательного прохода с проверкой подлинности найденных пиков на кратность их частот, с последующим выделением цепочек с наибольшей усредненной амплитудой, позволяет с большой точностью и малыми затратами времени находить форманты частотного среза. Последнее обстоятельство позволяет применять данный алгоритм при оперативной идентификации голосовых сообщений.

СПИСОК ЛИТЕРАТУРЫ

1. **Schroeder M.R.** Computer Speech: Recognition, Compression, Synthesis. http://datacompression.info/index.shtml
2. **Сердюков В.Д.** Опознавание речевых сигналов на фоне мешающих факторов.- Тбилиси: Наука, 1987. -142 с.
3. **Фланаган Дж.Л.** Анализ, синтез и восприятие речи. –М.: Связь, 1968. -396 с.
4. **Рабинер Л.Р., Шафер Р.В.** Цифровая обработка речевых сигналов. –М.: Радио и связь, 1981. - 496 с.
5. **Харари Ф.** Теория графов. – М.: Мир, 1973. -300 с.

Ереванский НИИ математических машин. Материал поступил в редакцию 10.02.2005.

Ա. Գ. ՀՈՎՀԱՆՆԻՍՅԱՆ

ՁԱՅՆԱՅԻՆ ԱԶԴԱՆՇԱՆԻ ՏԱՐՐԱՊԱՏԿԵՐԻՑ ՖՈՐՄԱՆՏՆԵՐԻ ԱՌԱՆՁՆԱՑՄԱՆ ԱՐԴՅՈՒՆԱՎԵՏ ԱԼԳՈՐԻԹՄ

Ներկայացված ալգորիթմն առանձնացնում է ձայնային ազդանշանի ֆորմանտները: Օգտագործելով տարրապատկերային վերլուծության միջոցները և ձայնային ազդանշանի ֆորմանտների հատկությունները՝ ալգորիթմն ապահովում է տոնային հաճախականության կապուղիներով փոխանցված ազդանշանի ֆորմանտների առանձնացման բարձր ճշգրտություն և արագություն: Ալգորիթմը կարող է օգտագործվել որպես գործիք՝ խոսողի նույնականացման համակարգերում:

Առանցքային բառեր. ֆորմանտ, ձայնային ազդանշան, տարրապատկեր, նույնականացում:

A.G. HOVHANNISYAN

EFFECTIVE VOICE SIGNAL FORMANT RECOGNITION ALGORITHM

The algorithm presented recognizes formants of voice signal. It uses spectrum analysis techniques and formant behavior characteristics to get high speed and reliability of recognition process. Voice signals in tone frequency channels in 300...3400 *Hz* range are given. The algorithm can be used as a tool in speaker authentication systems.

Keywords: formant, voice signal, spectrum, identification.