

Лабораторная работа №3

Методы оценивания частоты основного тона

Принято считать [2 – 4], что на участках вокализованного звука речевой тракт человека возбуждается периодическим колебанием связок. Период этого колебания называют периодом основного тона. Эта величина является индивидуальной характеристикой диктора. Она может меняться в зависимости от эмоциональной окраски речи, но в достаточно узких пределах. При параметрическом кодировании речи предполагают, что частота основного тона человека лежит в пределах 80 – 400 Гц.

Методы определения (выделения) основного тона можно разделить на следующие группы:

- амплитудная селекция;
- корреляционные методы;
- частотная селекция.

В данной работе рассматриваются некоторые алгоритмы для каждой из указанных групп.

1. Амплитудная селекция

На стационарном участке вокализованного звука при малом уровне шумов форма речевого колебания почти точно повторяется на каждом очередном периоде основного тона. Расстояние между максимумами максимумумами речевого сигнала можно приблизительно считать равными периоду основного тона. Основная трудность алгоритмов амплитудной селекции состоит в необходимости подавления локальных ложных максимумов. Этого можно добиться за счет повышения порога срабатывания в схеме поиска максимумов. Однако при этом увеличивается вероятность пропуска истинного максимума. Очевидно, что как пропуск, так и потеря максимума может привести к существенным искажениям

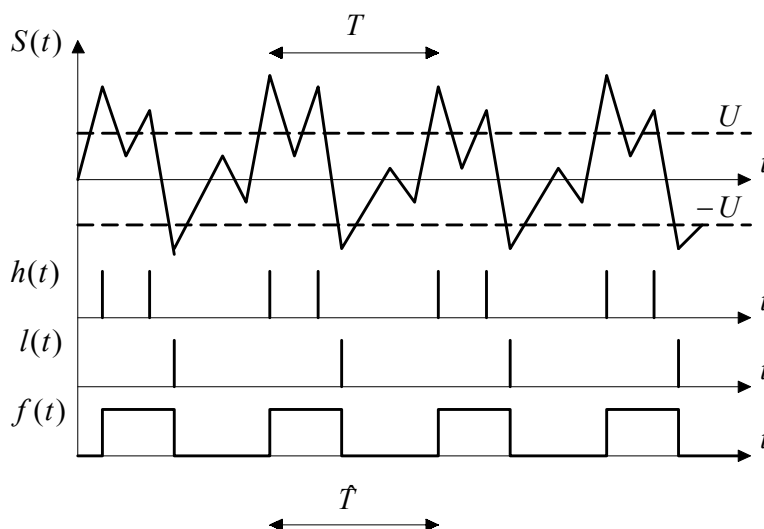


Рис. 3.1. Временная диаграмма работы устройства амплитудной селекции частоты основного тона

синтезированного звука. Повысить надежность определения периода основного тона можно, например, добавив второй канал амплитудной селекции, выделяющий положение минимумов речевого сигнала. На рис.3.1 приведены временные диаграммы работы такого устройства. Здесь $S(t)$ – речевой сигнал с периодом основного тона T . Сигналы $h(t)$ и $l(t)$ являются последовательностями импульсов на позициях, соответствующих максимальным и минимальным значениям речи. Эти импульсы управляют триггером, на выходе которого сформирован сигнал $f(t)$ с периодом \hat{T} , близким к T .

Главным достоинством устройств временной селекции является чрезвычайная простота реализации. Основной недостаток – невысокие точность и устойчивость определения основного тона.

2. Корреляционные методы определения периода основного тона

Пусть речевой сигнал представлен в виде последовательности отсчетов $S_i, i = 0, 1, 2, \dots$. Для вокализованных звуков можно считать, что

$$S_n \approx S_{n-T},$$

где T – период основного тона, выраженный в числе отсчетов. В качестве его оценки в момент времени n выберем значение k , минимизирующее функцию

$$L(n, k) = \sum_{i=0}^{N-1} (S_{n+i} - S_{n-k+i})^2. \quad (3.1)$$

Предположим, что энергия речевого сигнала не меняется на участке квазистационарности. Тогда оценка периода основного тона должна максимизировать корреляционную функцию

$$R(n, k) = \sum_{i=0}^{N-1} S_{n+i} S_{n-k+i}.$$

Данный подход обеспечивает существенно более высокую достоверность определения периода основного тона по сравнению с методами временной селекции. При этом следует отметить значительную вычислительную сложность данного алгоритма. Существуют его модификации, основанные на вычислении взаимной корреляционной функции

$$\tilde{R}(n, k) = \sum_{i=0}^{N-1} S_{n+i} F(S_{n-k+i}). \quad (3.2)$$

Подобрав удобную функцию $F(x)$, можно упростить алгоритм, сделав его пригодным для аппаратной реализации. Пусть функция $F(x)$ клиппирует речевой сигнал на три уровня: $\{-1, 0, 1\}$. Тогда вычислитель взаимной корреляционной функции (3.2) можно построить без умножителя.

Рассмотренные корреляционные методы оценивания периода основного тона имеют общий недостаток: неустойчивую работу в случае, когда речевой сигнал модулирован по амплитуде. Энергия же реальной, т.е. эмоционально окрашенной речи изменяется даже на квазистационарных участках, соответствующих одной фонеме. Модифицируем целевую функцию (3.1)

$$L(n, k) = \sum_{i=0}^{N-1} (S_{n+i} - \alpha_k S_{n-k+i})^2.$$

Параметр α_k имеет смысл коэффициента усиления. Легко показать, что для сдвига k оптимальное значение α_k вычисляется по формуле

$$\alpha_k = \frac{\sum_{i=0}^{N-1} S_{n+i} S_{n-k+i}}{\sum_{i=0}^{N-1} S_{n-k+i}^2}.$$

Тогда в качестве оценки периода основного тона в момент времени n следует выбрать такое значение k , которое максимизирует функцию

$$M(n, k) = \frac{\left(\sum_{i=0}^{N-1} S_{n+i} S_{n-k+i} \right)^2}{\sum_{i=0}^{N-1} S_{n-k+i}^2}.$$

Этот метод позволяет получить достаточно точную оценку основного тона, которая плавно меняется во времени в соответствии с изменениями голоса. Поэтому данный алгоритм используется в стандарте G.723, регламентирующем способ сжатия речевого сигнала для видеоконференций.

3. Частотная селекция

При вокализованном возбуждении речевого тракта в спектре сигнала присутствуют пики на частотах, кратных частоте основного тона. Если построить дискретное преобразование Фурье с достаточно малым шагом дискретизации по частоте, то можно попытаться в качестве оценки частоты основного тона использовать частоту, соответствующую максимальному значению энергии спектра. Поиск максимума следует производить в интервале 80 – 400 Гц. Однако часто возникает ситуация, когда в указанной полосе лежит и вторая гармоника основного тона, иногда даже с большей энергией. В этом случае она будет ошибочно принята за оценку основного тона. Чтобы избежать этого, будем искать максимум не спектра $X_n(k)$, а некоторой функции

$$P_n(k) = \prod_{r=1}^R |X_n(kr)|^2,$$

где индекс n указывает на то, что и спектр $X_n(k)$, и функция $P_n(k)$ вычислены в момент времени n . Учитывая то, что логарифм монотонно возрастает в области допустимых значений, целевая функция принимает вид

$$\tilde{P}_n(k) = \frac{1}{2} \ln(P_n(k)) = \sum_{r=1}^R \ln(|X_n(k \cdot r)|).$$

Эта функция представляет собой сумму R сжатых по частоте в r раз логарифмов спектра мощности. Суть идеи состоит в том, что для истинной частоты основного тона вторая гармоника второго слагаемого сложится с первой гармоникой первого слагаемого и усилит ее. Аналогично для третьего слагаемого и т. д. В результате для вокализованного звука будет иметь место ярко выраженный пик функции $\tilde{P}_n(k)$ на частоте основного тона. Для невокализованного звука суммирование будет иметь хаотический характер. Рис.3.2 иллюстрирует описанный метод.

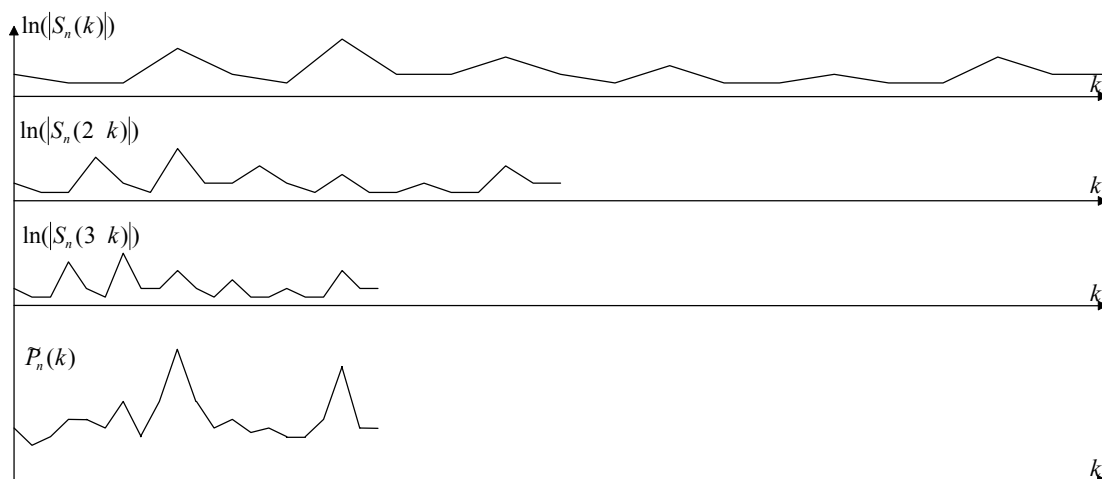


Рис. 3.2 . Иллюстрация метода частотной селекции периода основного тона

ЗАДАНИЯ

Исходные данные: файл, содержащий последовательность отсчетов речевого сигнала, тип алгоритма определения основного тона, интервал оценивания периода основного тона.

1. Разработать программу моделирования заданного алгоритма выделения периода основного тона.
2. Для файла, содержащего реальный речевой сигнал получить последовательность оценок периода основного тона и построить график его зависимости от номера интервала оценивания.
3. Визуально оценить период основного тона для каждого интервала, используя программу графического отображения речи.
4. Сравнить последовательность значений периода основного тона, полученных программой с последовательностью, полученной визуальным оцениванием.
5. Оценить диапазон значений периода основного тона, длину интервала, на которой период основного тона можно считать неизменным. Исходя из этих данных оценить величину битовых затрат на передачу информации об изменениях периода основного тона.