

- [1] Jones R.C. // J. Opt. Soc. Amer. 1941. Vol. 31. N 7. P. 488–493.
- [2] Jones R.C. // J. Opt. Soc. Amer. 1947. Vol. 37. N 2. P. 110–112.
- [3] Аззам Р., Башара Н. Эллипсометрия и поляризованный свет. М., 1981. 583 с.
- [4] Шерклифф У. Поляризованный свет. М., 1965. 246 с.
- [5] Laue M. // Ann. Phys. 1907. Vol. 23. N 1. P. 795–799.
- [6] Zernike F. // Physica. 1938. Vol. 5. P. 785–802.
- [7] Gabor D. // Proc. Symp. on Astr. Optics. Amsterdam, 1956. P. 17.
- [8] Борн М., Вольф Э. Основы оптики. М., 1970. 855 с.
- [9] Корн Г., Корн Т. Справочник по математике. М., 1973. 831 с.
- [10] Какичашвили Ш.Д. Поляризационная голография. Л., 1989. 142 с.
- [11] Какичашвили Ш.Д. // ЖТФ. 1989. Т. 59. Вып. 2. С. 26–34.
- [12] Какичашвили Ш.Д., Пурцеладзе А.Л. // Письма в ЖТФ. 1992. Т. 18. Вып. 22. С. 27–30.

08;12
© 1995 г.

Журнал технической физики, т. 65, в. 7, 1995

СПОСОБ ВЫДЕЛЕНИЯ ПЕРИОДОВ РЕЧЕВОГО СИГНАЛА

В.В.Дубровский, А.И.Егоров

Иркутский вычислительный центр СО РАН,
664033, Иркутск, Россия
(Поступило в Редакцию 13 июля 1994 г.)

В математической модели восприятия речи [1] исходная информация представлена энергетическими спектрами (ЭС) речевого сигнала (РС). При моделировании необходима также информация о средних значениях периодов РС на интервалах его анализа. В работе предлагается один из способов, обеспечивающих помехозащищенность и точность выделения периодов речи по ЭС.

В основу способа положен известный факт о кратности частот гармонических составляющих (линий) ЭС вокализованных элементов речи частоте основного тона. Следовательно, определив номер n какой-либо гармоники ЭС, можно рассчитать период сигнала $T = n/f_n$, где f_n — частота гармоники ЭС с номером n .

В общем случае в текущих ЭС речи находят отражение процессы, приводящие к резкому уменьшению интенсивности некоторых линий. В качестве критерия существования анализируемой линии в ЭС нами принято условие $L_I > L_n$, где L_I — уровень интенсивности линии; L_n — уровень обобщенного шума процедуры спектрального анализа. При выборе L_n учтены как погрешности спектрального анализа, так и уровень акустического шума. Для уровней РС порядка 65–85 дБ и акустического шума 50–55 дБ при принятом в модели способе получения ЭС (быстрое преобразование Фурье) уровень L_n составил около 60 дБ.

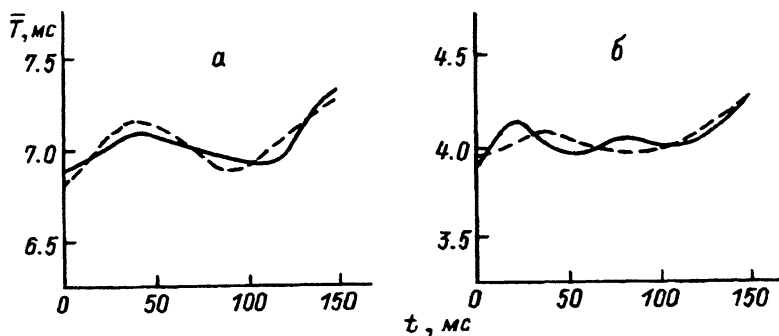
С целью повышения воспроизводимости выделения периодов РС в качестве линии (назовем ее опорной), номер которой необходимо определить, выбрана гармоника сигнала из частотного диапазона 90–360 Гц

с максимальным уровнем интенсивности. Согласно [2], приведенный диапазон частот с достоверностью ≥ 0.95 соответствует всем возможным значениям периодов в русской речи.

В способе производится декомпозиция диапазона 90–360 Гц на 3 частотные области, в каждой из которых для номера n опорной линии выполняется одно из условий: 1) $F \in [90, 180)$, $n = 1$; 2) $F \in [180, 270)$, $n \leq 2$; 3) $F \in [270, 360)$, $n \leq 3$, где F — частота опорной линии. В зависимости от того, в какую область частот попала опорная линия, предусмотрено 3 варианта поиска n . Отметим также, что в вариантах поиска 2 и 3 использовано свойство, выявленное при экспериментальных исследованиях озвученных элементов речи группы дикторов, состоявшей из 15 мужчин и 5 женщин. Это свойство заключается в следующем: при уровнях речи ≥ 65 дБ уровни интенсивности второй и как минимум одной из гармоник спектра с нечетным номером удовлетворяют вышеприведенному критерию существования линий в ЭС.

Далее рассмотрим варианты поиска n . В варианте 1 ($F \in [90, 180)$) значение $n = 1$ очевидно. В варианте 2 ($F \in [180, 270)$) производится поиск дополнительных к опорной линий в окрестностях частот $F/2$, $3F/2$, $5F/2$, ... Если найдена хотя бы одна линия, то $n = 2$. При отсутствии линий $n = 1$. Вариант 3 ($F \in [270, 360)$) заключается в поиске дополнительных линий в частотной области $F/3 - F$. Возможны такие результаты поиска: а) не найдено линий, б) найдены одна или две дополнительные линии. В случае реализации пункта а делается вывод о том, что $n \neq 3$ и вариант 3 приводится к варианту 2. Если найдены две линии, то $n = 3$. При поиске линий по варианту 3 наиболее часто в спектре содержится одна дополнительная линия, следовательно, n равно 2 или 3. В этой ситуации предусмотрена проверка на кратность между частотами опорной и дополнительной линий. Если частоты кратны, то $n = 2$. В противном случае $n = 3$.

Как уже отмечалось, при моделировании восприятия речи [1] используются средние значения периодов РС на интервалах анализа. Исходя из этого длительность временного окна, скользящего вдоль последовательности дискретных отсчетов РС, выбрана равной 30 мс. При такой длительности окна участок сигнала, по которому производилось выделение средних значений периодов РС, включал, как правило, не менее 3 периодов. Кроме того, за счет выбора относитель-



Траектории средних значений периодов речевого сигнала для мужского (а) и женского (б) голосов. Сплошная кривая — по предлагаемому способу, штриховая — расчет по результатам визуального выделения периодов.

но большой длительности окна и применения метода неравномерного взвешивания исходной последовательности окном Хэмминга [3] удалось повысить стабильность спектральных оценок для вокализованных элементов речи и получить приемлемое разрешение в частотной области. Использование же при расчете перекрывающихся во времени окон (шаг скачков выбран равным 1 мс) позволило сохранить в ЭС информацию о динамических свойствах РС.

С целью повышения точности выделения периодов РС применена эмпирическая процедура уточнения частоты линий в ЭС

$$f = f_i - \Delta f(I_{i-1} - I_{i+1}) / (I_{i-1} + I_i + I_{i+1}),$$

где f — уточненная частота линии в ЭС; f_i — частота спектральной составляющей с номером i , соответствующей локальному максимуму в огибающей ЭС; $\Delta f = 31.25$ Гц — частотное разрешение исходных составляющих ЭС; I_{i-1} , I_i , I_{i+1} — соответственно интенсивности составляющих ЭС с номерами $i-1$, i , $i+1$.

Качество выделения средних значений периодов РС по предлагаемому способу иллюстрируется рисунком. За эталон принята зависимость от времени средних периодов \bar{T} , рассчитанная на основе визуально выделенных значений периодов сигнала. Из рисунка видно, что погрешность выделения периодов РС не превышает 120 мкс. Испытание способа на большой группе дикторов ($\bar{T} = 3-10$ мс) позволило сделать вывод о том, что при более высокой экспрессности обеспечиваемые им помехозащищенность и точность выделения периодов РС находятся на уровне результатов, приведенных в [4].

Список литературы

- [1] Дубровский В.В., Егоров А.И. // Тр. Всероссийской школы "Компьютерная логика, алгебра и интеллектуальное управление". Иркутск. 1994. Т. 2.
- [2] Михайлов В.Г., Златоустова Л.В. Измерение параметров речи. М.: Радио и связь, 1987. 168 с.
- [3] Рабинер Л., Гоулд Б. Теория и применение цифровой обработки сигналов. М.: Мир, 1978. 848 с.
- [4] А.С. № 1583970 СССР. БИ № 29. 1990.