

О. КЮННАП

ПРИМЕНЕНИЕ ДИСКРЕТНОГО ПРЕОБРАЗОВАНИЯ ЛАПЛАСА К АЛГОРИТМУ ОБРАТНОЙ ФИЛЬТРАЦИИ ДЛЯ ОПРЕДЕЛЕНИЯ ФОРМАНТ РЕЧЕВОГО СИГНАЛА

O. KUNNAP. DISKREETSE LAPLACE'I TEISENDUSE RAKENDAMINE KONESIGNAALI FORMAN-
 TIDE ERALDAMISE PÕRDFILTRI ALGORITMILE

O. KUNNAP. THE DISCRETE LAPLACE TRANSFORM APPLICATION TO THE SPEECH FORMANT
 TRAJECTORY ESTIMATION INVERSE FILTER ALGORITHM

Введение. Согласно общепринятой теории речеобразования, речь является результатом фильтрации сигнала голосового источника голосовым трактом. Передаточную функцию голосового тракта можно рассматривать состоящей только из полюсов. Значения этих полюсов изменяются во времени сравнительно медленно, что позволяет считать их усредненные значения на протяжении 10—20 мсек постоянными. В частотной области полюс выявляется резонансным максимумом, называемым формантой.

За последнее время для определения параметров голосового тракта разработан целый ряд методов. Наилучшие результаты дает метод линейного прогнозирования. Вариант этого метода, приведенный Дж. Д. Маркелом [1], взят за основу при определении коэффициентов a_h передаточной функции $H(z)$ голосового тракта и в данной работе.

Определение коэффициентов передаточной функции голосового тракта. Параметры голосового тракта находят, минимизируя энергию выходного сигнала цифрового фильтра

$$\hat{H}(z) = 1 + \sum_{h=1}^p \hat{a}_h z^{-h} \quad (1)$$

методом наименьших квадратов. Модельный фильтр $\hat{H}(z)$ является обратным к фильтру голосового тракта $H(z)$. Энергия Q выходного сигнала фильтра $\hat{H}(z)$ выражается в виде

$$Q = \sum_{n=1}^N (x_n + \sum_{h=1}^p \hat{a}_h x_{n-h})^2, \quad (2)$$

где x_n , $n = 1, 2, \dots, N$; $x_n = 0$; $n \leq 0$ — выборки речевого сигнала. Приравняв частные производные $\partial Q / \partial \hat{a}_h$ к нулю и решая уравнения относительно \hat{a}_h , получаем систему уравнений

$$R \times \bar{a} = -\bar{c}, \quad (3)$$

где R — матрица порядка $p \times p$ с элементами $r_{i,j} = c_{|i-j|}$, а c_m — коэффициенты автокорреляции сигнала, определяемые по формуле

$$c_m = \sum_{n=1}^{N-m} x_n x_{n+m}, \quad m = 0, 1, \dots, p. \quad (4)$$

Элементы вектора $\bar{a} = (1, \hat{a}_1, \dots, \hat{a}_p)^T$ являются искомыми, элементы вектора $\bar{c} = (c_1, c_2, \dots, c_p)^T$ находятся из формулы (4). Решая систему уравнений (3), получаем коэффициенты \hat{a}_k , которые являются наилучшими в смысле наименьшей квадратичной ошибки к выбранной обратной модели речевого тракта.

Определение формантных частот. В качестве формантных частот можно выбрать частоты полюсов с меньшей вещественной частью на s -плоскости. Значения полюсов можно получить, решая полиномиальное уравнение с коэффициентами \hat{a}_k . Менее трудоемким является способ, при котором вычисляется амплитудно-частотная характеристика передаточной функции $1/\hat{H}(z)$ и выбираются частоты наиболее выраженных максимумов в качестве формантных частот [1]. Однако при близком расположении двух полюсов их резонансы в спектре сливаются в один максимум, а полюса, отдаленные от мнимой оси s -плоскости, имеют слабовыраженные максимумы. Этих недостатков можно избежать, вычисляя спектр по контуру, установленному ближе к полюсам на s -плоскости. Используемый для этой цели алгоритм z -преобразования [2] весьма трудоемок. Применение же дискретного преобразования Лапласа (ДПЛ) позволяет получить сравнимый эффект при очень малом числе вычислений.

ДПЛ является частным случаем z -преобразования

$$X_k = \sum_{n=0}^{N-1} s_n z^{-n}, \quad k=0, 1, \dots, N-1, \quad (5)$$

где $s_n, n=0, 1, \dots, N-1$ — выборки сигнала. Заменяя $z = e^{-(\sigma + j2\pi k/N)T}$, где T — период измерения сигнала, получаем выражение ДПЛ

$$X_k = \sum_{n=0}^{N-1} s_n e^{-(\sigma + j2\pi k/N)Tn} = \sum_{n=0}^{N-1} (s_n e^{-\sigma Tn}) e^{-j2\pi kTn/N}, \quad (6)$$

$$k=0, 1, \dots, N-1,$$

откуда видно, что ДПЛ можно вычислять дискретным преобразованием Фурье от последовательности $s_n, n=0, 1, \dots, N-1$, умноженной на весовую функцию

$$\omega_n = e^{-\sigma Tn}, \quad n=0, 1, \dots, N-1, \quad (7)$$

где σ определяет удаление оси интегрирования от мнимой оси s -плоскости.

Экспериментальные результаты. Исследуемый речевой сигнал измерялся с частотой 10 кГц и дифференцировался для поднятия высших частот. Число коэффициентов $\hat{a}_k, k=1, 2, \dots, p$, выбрано равным 15. Речевой сигнал взвешивался окном Гамминга длиной в 25,5 мсек.

На рис. 1 показан спектр сегмента гласной [I] в слове «resorts», вычисленный от коэффициентов \hat{a}_k преобразованием Фурье (а) и преобразованием Лапласа по контуру 100 гц левее мнимой оси s -плоскости (б). Как видно, в спектре на рис. 1, б выявляется форманта с частотой 1836 гц, а на рис. 1, а она отсутствует.

Полюса речевого тракта обычно лежат в пределах от 25 до 150 гц левее мнимой оси s -плоскости, что позволяет располагать ось интегрирования в этих пределах. Опытным путем выбран контур —100 гц, так как при этом выявляются наиболее близко расположенные форманты, но не теряются и те, которые находятся ближе к мнимой оси.

На рис. 2, а показана картина видимой речи фразы: «Resorts vanishing forever», произнесенной диктором мужчиной, на рис. 2, б и 2, в —

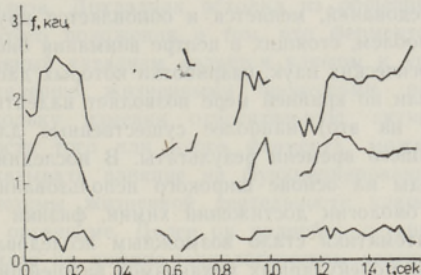
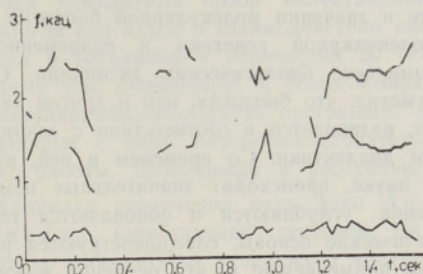
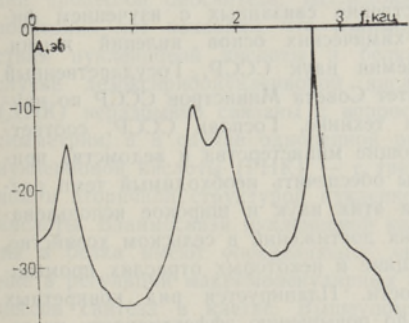
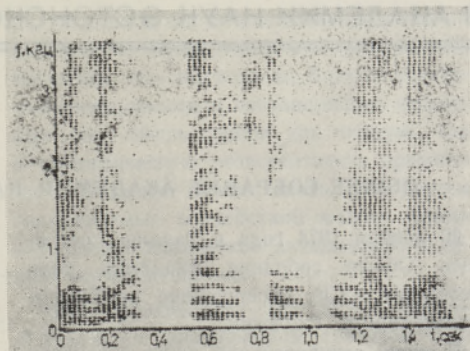
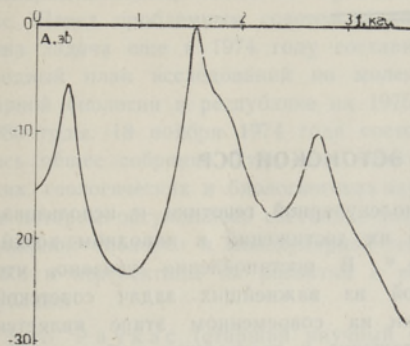


Рис. 1. Спектральные срезы одного сегмента гласной [I], вычисленные по контурам $\sigma=0$ (а) и $\sigma=-100$ гц (б).
Рис. 2. Анализ фразы: «Resorts vanishing forever». Спектрограмма (а) и траектории формант, вычисленные по контурам $\sigma=0$ (б) и $\sigma=-100$ гц (в).

траектории формант на озвученных участках речи этой же фразы, вычисленные по контурам 0 и 100 гц левее мнимой оси s -плоскости. Озвученные участки речи выбраны по энергии сигнала и по энергии высших частот в спектре. На рис. 2, б можно заметить несколько пробелов в траекториях формант, заполненных на рис. 2, в.

Закключение. Изложенный выше метод позволяет получать качественные траектории формант речевого сигнала без больших затрат машинного времени. Вычисление спектров методом ДПЛ требует небольшого числа операций, но значительно улучшает качество спектральных срезов, благодаря чему не требуется применения komplицированных алгоритмов выбора трех формант. При использовании ДПЛ тремя формантами являются три первых резонансных максимума спектральных срезов.

ЛИТЕРАТУРА

1. Markel J. D., IEEE Trans. Audio Electroacoust., AU-20, 129—137 (1972).
2. Rabiner L. R., Schafer R. W., Rader C. M., IEEE Trans. Audio Electroacoust., AU-17, 86—92 (1969).