ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ
1971 г. вып. 44

УДК 621.391:681.3.06.

АНАЛИЗ И СИНТЕЗ РЕЧИ НА ОСНОВЕ Z-ОПИСАНИЯ

В.С. Лозовский

I. Аппроксимация

В работе [1] предложено для звуков речи с тональным возбуждением аппроксимировать отклик речеобразующего тракта на участке смыкания голосовых связок дробно-рациональной функцией вида :

$$\hat{H}(z) = \frac{N(z)}{D(z)} = \frac{a_0 + a_1 z^{-1} + \dots + a_n z^{-n}}{1 + b_1 z^{-1} + \dots + b_m z^{-m}}, \qquad (1)$$

где $z = e^{\tau_0}$. Выражение (1) представляет собой передаточную функцию дискретной линейной динамической системы, полюса которой соответствуют формантам исследуемого речевого сигнала. При этом не выдвигается требование наблюдаемости всех состояний речеобразующей системы, поскольку для слухового восприятия важны не собственные частоты реальной физической системы, а лишь те колебания, которые присутствуют в конкретном, воспринимаемом ухом сигнале.

Для сегментов речи с шумовым возбуждением предлагается воспользоваться методами идентификации линейных динамических систем в стохастическом случае [2]. Итак, в каждый дискретный момент времени t сигнал на выходе аппроксимирующей системы, описываемой выражением (1), имеет вид :

$$y_t = -b_1 y_{t-1} - \dots - b_m y_{t-m} + a_0 x_t + \dots + a_n x_{t-n}, \qquad (2)$$

где x_t и y_t – отсчеты соответственно входного и выходного сигналов.

Предполагается, что при произнесении фрикативных звуков на вход резонансной системы речеобразующего тракта подан белый шум. В модели ему соответствует дискретный белый шум – сигнал с некоррелированными соседними отсчетами, значения которых распределены равномерно, например, в интервале (1, -1).

Рассмотрим систему, описываемую уравнением 1/ $D(z)$, полагая, что на её вход поступает сигнал :

$$u_t = \sum_{j=0}^{r} a_j x_{t-j}. \qquad (3)$$

Множество $\{u_t\}$ представляет собой так называемый линейный шум, в котором соседние отсчеты коррелированы. Из (3) следует, что с увеличением интервала между рассматриваемыми значениями u_t корреляция между ними уменьшается; отсчеты, отстоящие друг от друга на интервал, не меньший $r+1$, некоррелированы.

Перепишем (2) в виде:

$$-\sum_{j=1}^{m} b_j y_{t-j} = y_t - u_t \triangleq f_t. \qquad (4)$$

Итак, f_t – это значения функции, аппроксимации которой в наилучшем смысле мы будем добиваться, выбирая b_j. Функция $\{f_t\}$ включает в себя наблюдаемую составляющую y_t и линейно-шумовую составляющую $-u_t$, нам не известную. Выбрав отсчеты f_t с шагом $r+1$, мы переходим от линейного к белому шуму.

Составим систему из m уравнений (4), полагая $f_t = y_t$. Восстановив по полученным b_j аппроксимирующую функцию, обнаруживаем, что на каждом уравнении мы допустили ошибку u_t, которая есть несмещенный белый шум. При большем числе исходных уравнений к ошибкам $|u_t|$ будут добавляться ошибки среднеквадратической аппроксимации, которые также носят характер белого шума. Таким образом, можно полагать, что полученная описанным путем оценка b_j будет несмещенной.

Подчеркнем, что в случае стохастической аппроксимации определяются лишь коэффициенты $D(z)$.

При решении задач z-идентификации, о которых шла речь выше, возникает необходимость в решении переопределенных систем вида :

$$-\sum_{j=1}^{n} b_j y_{t-j} = y_t, \quad t = 1, \dots, m; \quad m \geq n. \qquad (5)$$

Можно показать, что подобные системы являются плохо обусловленными; поэтому корректный выбор вычислительной процедуры для их решения является весьма важной задачей. Использование известной трансформации Гаусса позволяет преобразовать матрицу системы (5) к квадратной симметричной положительно-определенной форме; для решения полученной системы весьма эффективен метод квадратных корней. Однако точность полученного решения оказывается недостаточной. Дело в том, что уже на этапе составления нормальной системы при вычислении скалярных произведений допускаются ошибки округления, которые в условиях слабой определенности (5) иногда приводят даже к отрицательно определенным системам. Конечно, можно вычислять скалярное произведение и даже полностью решать полученную систему с удвоенным числом знаков мантиссы, но, как показал эксперимент, в рассматриваемом случае этого недостаточно. Хорошие результаты дает использование матриц отражения Хаусхолдера [3] для триангуляризации исходной системы (5); этот алгоритм с некоторыми модификациями и был принят в качестве рабочего.

Обсуждая методы решения переопределенных систем и делая определенные выводы относительно физической стороны исследуемых процессов, нельзя обойти вопрос о корректности решаемой задачи. В работе [4] показано, что при задании исходной информации (значения y_t, y_{t-j} в (5)) с некоторым приближением задача определения нормального решения системы (5) некорректна в смысле Адамара. Некорректность проявляется в больших вариациях решений b_j при малых вариациях исходных данных. Прежде чем приступить к регуляризации исходной системы, которая позволяет до некоторой степени "выправить" решение, остановимся на физическом смысле некорректности нашей задачи.

Итак, нас интересует построение модельной системы, отклик которой с нужной точностью аппроксимирует анализируемый речевой сигнал. Ясно, что существует множество систем, отвечающих этому требованию. Разница между откликами этих систем и анализируемым сигналом, поскольку мы используем методы среднеквадратической аппроксимации, носит характер белого шума. Некорректность нашей задачи проявляет себя в том, что близкие в указанном смысле во временной области системы могут заметно отличаться по положению полюсов, т.е. формант. Это приводит к выводу, что, с одной стороны, при вычислении положения формант для соседних сегментов речи возможны их резкие скачки, а, с другой стороны, эти нестационарности отражают физику явления в рамках рассматриваемой модели и на качестве синтеза по полученным параметрам сказываться не должны. Здесь уместно снова подчеркнуть, что наша задача – построить модельную систему, близкую исследуемой во временной области, а не определить передаточную функцию речеобразующего тракта. По-видимому, вторая задача не может быть строго решена без получения более подробной информации об источнике возбуждения [*], чем мы можем почерпнуть из исследования лишь осциллограммы речи.

Теперь о регуляризации. Потребность в ней может возникнуть, например, из соображений уменьшения ширины полосы сигнала при передаче значений коэффициентов аппроксимации для синтеза речи на приемном конце. Суть известных методов регуляризации задач линейной алгебры [4] сводится к тому, что минимизируемый функционал (в нашем случае сумма квадратов ошибок аппроксимации) дополняется функционалом, зависящим от разности между искомым решением и некоторым фиксированным вектором, например, нулевым, или же, в нашем случае, равным вектору неизвестных коэффициентов, определенному для предыдущего сегмента.

Выведем необходимые соотношения. Пусть исходная система уравнений имеет вид :

$$a_{1:m,1:n} \times z_{1:n} = b_{1:m}, \quad m \geq n. \qquad (6)$$

Здесь одна индексная пара соответствует вектору-столбцу, а две – матрице. Минимизируемый функционал имеет вид :

$$F = \| a_{1:m,1:n} \times z_{1:n} - b_{1:m} \|^2 + \alpha \| z_{1:n} - z^0_{1:n} \|^2 = \sum_{i=1}^{m} \Big(\sum_{j=1}^{n} a_{ij} z_j - b_i \Big)^2 + \alpha \sum_{j=1}^{n} (z_j - z_j^0)^2, \qquad (7)$$

где $z^0_{1:n}$ – некоторый фиксированный вектор, а α – параметр регуляризации.

$$\frac{\partial F}{\partial z_k} = 2 \sum_{i=1}^{m} \Big(\sum_{j=1}^{n} a_{ij} z_j - b_i \Big) a_{ik} + 2\alpha (z_k - z_k^0), \quad k = 1, \ldots, n.$$

Полученная квадратичная форма имеет лишь один экстремум :

[*] Строго говоря, при этом надо ещё потребовать наблюдаемости всех состояний тракта.

это – минимум, поскольку форма существенно положительна.

$$\sum_{j=1}^{n}\sum_{i=1}^{m} a^{\tau}_{ki}\, a_{ij}\, z_j + \alpha z_k = \sum_{i=1}^{m} a^{\tau}_{ki}\, b_i + \alpha z^{\circ}_k, \qquad k = 1, \ldots, n. \qquad (8)$$

Выражение (8) – регуляризованная система нормальных уравнений для решения задачи среднеквадратической аппроксимации. Большей наглядностью обладает матричная форма записи :

$$(a^{\tau}_{1:n,\,1:m} \times a_{1:m,\,1:n} + \alpha\, J_{1:n,\,1:n}) z_{1:n} = a^{\tau}_{1:n,\,1:m} \times b_{1:m} + \alpha z^{\circ}_{1:n} \qquad (9)$$

Здесь J – единичная матрица.

В связи с тем, что принятый нами алгоритм позволяет обойти составление нормальных уравнений, перейдем к записи исходной переопределенной системы. Предварительно перепишем (9) в блочном виде :

$$(a^{\tau}_{1:n,\,1:m} \mid \sqrt{\alpha}\, J_{1:n,\,1:n}) \begin{pmatrix} a_{1:m,\,1:n} \\ \sqrt{\alpha}\, J_{1:n,\,1:n} \end{pmatrix} z_{1:n} = (a^{\tau}_{1:n,\,1:m} \mid \sqrt{\alpha}\, J_{1:n,\,1:n}) \begin{pmatrix} b_{1:m} \\ \sqrt{\alpha}\, z^{\circ}_{1:n} \end{pmatrix} \qquad (9')$$

Отсюда легко получить исходную систему :

$$\begin{pmatrix} a_{1:m,\,1:n} \\ \sqrt{\alpha}\, J_{1:n,\,1:n} \end{pmatrix} z_{1:n} = \begin{pmatrix} b_{1:m} \\ \sqrt{\alpha}\, z^{\circ}_{1:n} \end{pmatrix}, \qquad (10)$$

которая решается с использованием триангуляризации Хаусхолдера.

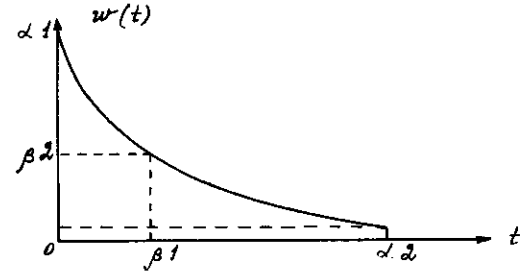
Дальнейшего повышения точности решения задачи z -идентификации можно добиться с помощью весовых функций. Весовая функция – это массив коэффициентов $w_{1:m+n}$, на которые должны умножаться уравнения исходной системы в соответствии с достоверностью каждого. Значения последних n весов уже определены – $\sqrt{\alpha}$; на выборе же первых m коэффициентов мы сейчас остановимся.

Прежде всего, трудно рассчитывать на то, что в реальном речевом сигнале при анализе озвученных сегментов удастся четко фиксировать границы интервала смыкания голосовых связок. Поэтому первым и последним уравнениям (10) (без учета регуляризующей добавки) следует придавать меньший вес. Подобное подавление граничных уравнений с помощью гладкой весовой функции будет также способствовать сглаживанию вектора решений при выполнении аппроксимации в скользящем режиме при поиске интервала аппроксимации. Специфика нашей задачи заключается в том, что аппроксимирующие функции вычисляются рекуррентно. При этом в связи с конечной точностью вычислений происходит накопление ошибок. Поэтому к концу интервала исходным уравнениям следует придавать меньший вес.

Практически использовались функции четырех типов :

Тип 1. Значение функции принимается равным постоянной величине на всем интервале аппроксимации.

Тип 2. Экспонента.



При этом задаются:

$\alpha 1$ – значение ординаты в начале,

$\beta 1$ и $\beta 2$ – абсцисса и ордината опорной точки,

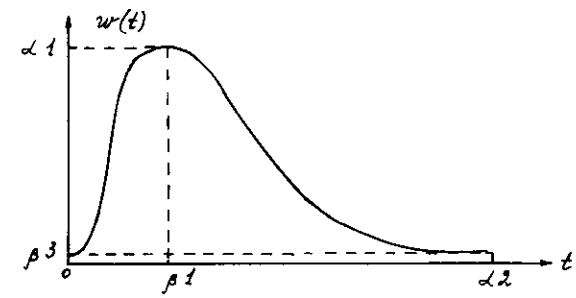
$\beta 3$ – ордината на конце интервала аппроксимации: $\alpha 2$.

Аналитически функция записывается в виде :

$$w(t) = \alpha 1 \cdot e^{\lambda 2 \left(\frac{t}{\lambda 1}\right)^{\lambda 1}}, \qquad (11)$$

где необходимые коэффициенты $\lambda 1$ и $\lambda 2$ легко определяются.

Тип 3. Сдвинутая косинусоида.



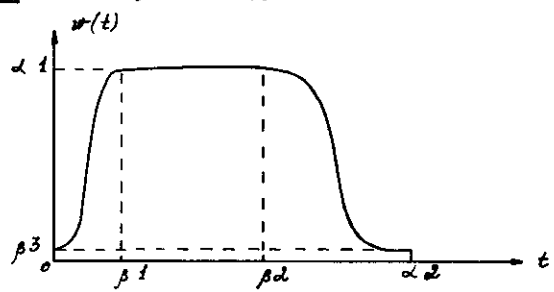
Функция имеет вид:

$$w(t)=\frac{1+\beta 3-\cos\left(2\pi\left(\frac{t}{2\alpha 2}\right)^{\lambda 1}\right)(1-\beta 3)}{2}\cdot\alpha 1, \quad (I2)$$

$$\lambda 1=\ln 0.5/\ln \beta 1,$$

$\beta 1$ и $\beta 3$ нормированы к единице.

Тип.4. Комбинированная функция.



Здесь передний фронт строится по формуле:

$$w_{пр}(t)=\frac{1+\beta 3-\cos\left(\frac{\pi t}{\beta 1\cdot\alpha 2}\right)(1-\beta 3)}{2}\alpha 1, \quad (I3)$$

а задний:

$$w_{зр}(t)=\frac{1+\beta 3+\cos\left(\frac{\pi(t-\beta 2\cdot\alpha 2)}{(1-\beta 2)\cdot\alpha 2}\right)(1-\beta 3)}{2}\cdot\alpha 1; \quad (I4)$$

на интервале ($\beta 1$, $\beta 2$) значение весовой функции постоянно и равно $\alpha 1$.

Варьируя параметр $\alpha 1$ весовых функции, удобно масштабировать исходную систему, что способствует дополнительному повышению точности вычислений.

2. Поиск интервала

Для озвученных сегментов речи не меньшее, чем сама аппроксимация, значение имеет поиск интервала, соответствующего смыканию голосовой щели, т.е. свободным колебаниям изучаемой системы. При этом попутно с поиском интервала можно было бы определять и период основного тона (ОТ), но проще это сделать отдельно.

Итак, исследуемый отрезок речи разбивается на сегменты по 20 мсек. Если энергия сигнала на сегменте меньше пороговой, аппроксимация не выполняется и сегменту присваивается код : П - пауза.

Для поиска ОТ на сегментах используется модификация авто - корреляционного метода [5] , в результате применения которого вычисляется индекс апериодичности сегмента (ИА). Для строго периодического сигнала ИА = 0. Устанавливается величина порога (~0,5). Считается, что при бо́льших значениях ИА мы имеем дело с шумовым возбуждением (код типа сегмента: Ш), при меньших - с тональным (код: Т). В последнем случае вычисляется частота ОТ. Для Ш - сегментов сразу выполняется аппроксимация сигнала, пропущенного через фильтр высших частот, по способу , описанному в разделе I. В Т - сегментах предварительно ищется оптимальный интервал аппроксимации.

Поиск выполняется методом пробной аппроксимации в скользящем режиме. Предварительно сигнал пропускается через фильтр низших частот *) для понижения частоты квантования вдвое. Этот прием облегчает работу алгоритма аппроксимации и сокращает машинное время. В качестве начальных точек для пробной аппроксимации выбираются не все точки, а лишь соответствующие моментам перехода сигнала через нуль, а также дополнительные точки с тем, чтобы расстояние между соседними не превышало задаваемого фиксированного порога.

При исследовании методов поиска было испробовано несколько алгоритмов.

И н т е р в а л п о с т о я н с т в а к о э ф ф и ц и е н т о в . Длина интервала при пробной аппроксимации выбирается заведомо меньше ожидаемой. Производится аппроксимация для выбранной последовательности начальных точек и сравниваются значения коэффициентов b_j (I), полученные для каждого участ-

*) Параметры рекурсивных ФНЧ и ФВЧ определялись по программе [6] ; на частоте квантования 16 кгц их данные:
ФНЧ 0 - 3700 гц , аппроксимация по Чебышеву пятого порядка, неравномерность на вершине : 0.04.
ФВЧ моделируется полосовым фильтром: $f_{ср}$ = 4000 гц, $(2\Delta f)_{0.707}$ = 6000 гц, аппроксимация по Чебышеву, порядок НЧ прототипа: 5 , неравномерность: 0.04.

ка. При закрытой голосовой щели система совершает свободные колебания, параметры её должны изменяться мало; с открытием щели меняются резонансные характеристики (главным образом, добротность), в систему поступает энергия извне, и параметры аппроксимации должны проявить заметную нестационарность. На практике же ситуация осложняется. Как уже говорилось в разделе I, решаемая задача анализа является некорректной, т.е. даже в условиях фактического постоянства физических параметров исследуемой системы им будут соответствовать заметно различающиеся значения параметров аппроксимации. Конечно, можно регуляризовать задачу, но и в этом случае нужно алгоритмически сформулировать определение, что считать стационарным участком, а что нет. Был исследован метод поиска интервала постоянства с использованием таксономии на потенциальных функциях, но сложность алгоритма и время вычислений не оправдывались качеством полученных результатов.

О б р а т н а я ф и л ь т р а ц и я . Сигнал ОТ можно выделить в чистом виде, если пропустить речь через фильтр, нули которого в точности соответствуют полюсам речевого тракта. Здесь получается заколдованный круг : ведь мы как раз ищем участок аппроксимации, чтобы по нему определить резонансные параметры тракта. Применение рассматриваемого метода упирается в выбор функционала оценки качества обратной фильтрации. А это не просто и требует дополнительных затрат машинного времени при вычислениях.

В з в е ш е н н а я о ш и б к а а п п р о к с и м а ц и и и э к с т р а п о л я ц и и . Этот метод требует небольших дополнительных затрат машинного времени, обладает достаточной надежностью, в связи с чем он и был принят в качестве рабочего *). Суть его в следующем.

Для аппроксимации реального сигнала выбрана линейная динамическая модель. При этом качество аппроксимации будет тем выше, чем точнее выполняются требования линейности для исследуемого сигнала. Кроме этого, требуется, чтобы аппроксимирующая

*) Первоначально [7] в качестве критерия использовался минимум ошибки аппроксимации, являющийся частным случаем описываемого здесь алгоритма.

система была пассивной. Таким образом, выбирая участок, на котором ошибка аппроксимации минимальна, мы автоматически получаем участок, который с наибольшим основанием может быть принят в качестве отклика линейной динамической системы. Точность определения параметров системы, а также неизменность их во времени должны подтверждаться небольшим значением ошибки экстраполяции в течение нескольких отсчетов сигнала, примыкающих к концу исследуемого интервала.

Практически описываемый алгоритм реализуется следующим образом. Выбирается для взвешивания исходной системы уравнений функция типа 3 (раздел I) с малым значением $\beta 1$. Производится определение коэффициентов b_j и уточнение a_j по всей длине рассматриваемого интервала [I] . В связи с малой величиной $\beta 1$ значение весовой функции $w(t)$ к концу интервала весьма невелико, т.е. последние уравнения исходной переопределенной системы практически уже не участвуют в определении коэффициентов. Таким образом, к концу интервала $\alpha 2$ уже имеет смысл говорить не об аппроксимации, а об экстраполяции исследуемого сигнала. Ошибка вычисляется как разность между исходной и аппроксимирующей функциями на всем интервале $0 - \alpha 2$ и без взвешивания нормируется по интенсивности сигнала. Запоминается значение этой ошибки для всех исследованных на данном сегменте интервалов и находится её минимум. Найденная точка считается началом оптимального интервала для точной аппроксимации. Степень аппроксимирующей функции определяется естественным образом: исходная система уравнений строится для максимальной степени, представляющей интерес, например, восьмой. В процессе триангуляризации системы контролируется норма ведущего столбца. Когда она станет меньше заданного порога (т.е. преобразующая матрица отражений вырождается в единичную), процесс обрывается и действительная степень аппроксимации полагается равной числу преобразованных столбцов.

В самом алгоритме аппроксимации не заложено требование пассивности. Поэтому после того как будет выполнена аппроксимация сигнала на выбранном участке, вычисляется [8] положение полюсов аппроксимирующего полинома. Если хотя бы один полюс выходит за пределы единичной окружности в z-плоскости, выбранный интервал считается активным, а аппроксимация повторяется для следующего по значению ошибки интервала аппроксимации.

3. О некоторых аспектах Z - анализа

Результатом решения задачи идентификации параметров речевого сигнала методами рациональной z - аппроксимации является представление вида $N(z)/D(z)$ для озвученных 20 мсек - сегментов и $a_0/D(z)$ - для шумовых. Вопрос об использовании $N(z)$ в первом случае не вполне ясен, в частности, потому, что не всегда в передаточной функции тракта нули отсутствуют, а отличить нули источника от нулей тракта пока не всегда удается ; с другой стороны, даже выявив нули источника, мы таким образом получим информацию лишь о конечном участке импульса возбуждения. Таким образом, остановимся пока на том, что сегменты обоих типов будут характеризоваться лишь полиномами $D(z)$(I).

Корни полинома $D(z)$ соответствуют формантам [I] ,т.е. в принципе отсюда можно определить и частоты, и ширину полос найденных формант. Возникает вопрос, удобно ли пользоваться формантным описанием, в частности, для передачи по каналам связи. В этой связи следует заметить, что до сих пор не известен строгий и однозначный алгоритм соотнесения спектральных максимумов (полюсов) и формант [9] . Если в определенной ситуации и удается найти первые четыре форманты, то через несколько десятков миллисекунд какая-либо из них может исчезнуть под влиянием нуля источника возбуждения, форманты могут слиться,могут появиться субформанты, назальные, ложные и т.п. При идентификации формант, например, по картинкам "видимой речи" рекомендуется [9] иметь в виду "критерий непрерывности". Однако этот критерий весьма проблематичен, поскольку некоторые полости при артикуляции открываются, закрываются, делятся на две и т.д.Большая часть имеющихся в литературе критериев рассчитана на то , что почти все фонемы уже распознаны, расчленены, изучены переходы; только после этого, опираясь на физическую модель тракта, можно пытаться соотнести найденные максимумы с номерами соответствующих формант.

В инженерной практике при построении формантных систем частотный диапазон делится на фиксированные, реже перестраивае - мые области, которые и соотносятся с формантами. Возможны выходы формант за границы выделенных областей. Ясно, что первый путь очень неформален, а второй излишне формализован,и построить корректную автоматическую схему формантного анализа по

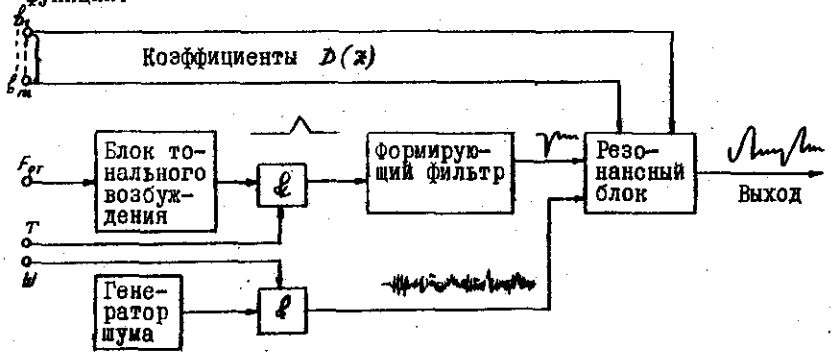
сигналу от микрофона на их основе очень непросто.

С другой стороны, методы z - анализа, описываемые в настоящей работе, будучи по существу описанием, близким к формантному, позволяют избежать нетривиальной задачи индексации формант при анализе путем передачи по каналу либо коэффициентов $D(z)$ непосредственно, либо подвергнув их предварительно-ортогональной кодировке. Это дает возможность избежать определенных затруднений при конструировании синтезатора (параллельная или последовательная схема, управление шириной полос , амплитуд формант, частотная коррекция и т.п.). В нашем случае синтезатор по коэффициентам $D(z)$ непосредственно во временной области восстанавливает исходный сигнал традиционными методами рекурсивной фильтрации.

4. z - синтез

Ниже приводится блок-схема предлагаемого синтезатора.В качестве управляющих используются следующие сигналы:

- тональное (Т) или шумовое (Ш) возбуждение.
- значение частоты основного тона $F_{от}$.
- значения m коэффициентов $D(z)$аппроксимирующей функции.



Блок тонального возбуждения генерирует треугольные импульсы. Назначение формирующего фильтра - придать возбуждающему импульсу форму, близкую к реальной. Критерий настройки - естественность синтезируемой речи. Фильтр включает от нуля до пяти резонаторов, которые могут использоваться в качестве полюсов и нулей. Резонансный блок реализует алгоритм рекурсивной фильтра-

ции :

$$y_t = x_t - b_1 y_{t-1} - \dots - b_m y_{t-m}, \qquad (15)$$

где x_t и y_t – отсчеты функций на его входе и выходе соответственно, а $b_1, \dots, b_m$ – коэффициенты аппроксимации $D(z)$.

Генератор шума выдает последовательность псевдослучайных чисел, равномерно распределенных в диапазоне (-0,5 , 0,5).

Простота схемы синтезатора избавляет от необходимости описывать принцип его работы.

5. Результаты эксперимента

Работа анализатора и синтезатора моделировалась на ЭВМ БЭСМ-6. С помощью специальной программы речевой сигнал вводился в машину через девятиразрядный аналого-цифровой преобразователь. Программа фиксировала начало и конец слова или фразы и переписывала сигнал в цифровом виде на ленту БЭСМ-6 для последующей обработки.

На рис. I иллюстрируется работа алгоритма z – аппроксимации. Основные параметры программы анализа:

Порядок $D(z)$ при поиске интервала аппроксимации : $m_1 = 6$.

При поиске коэффициенты b_j не уточнялись, а a_j определялись по всему интервалу.

Длина пробного интервала аппроксимации при поиске оптимального интервала $la_S = 0{,}9\ T_{or}$.

При поиске интервала исследуется участок сигнала длиной $l_S = 2.1\ T_{or}$.

Порог по нормам столбцов матрицы при триангуляризации $\varepsilon_S = 1_{10}{-}6$.

Порядок $D(z)$ при точной аппроксимации $m = 8$.

Осуществлялись две итерации уточнения b_j.

Длина интервала аппроксимации $l_a = 0{,}6\ T_{or}$.

Применена регуляризация: в систему включен модуль разности между вектором неизвестных коэффициентов и вектором коэффициентов $D(z)$, полученных на предыдущем сегменте. Параметр регуляризации $\alpha = 0.06$.

Весовая функция при поиске начала аппроксимации: Тип 3;

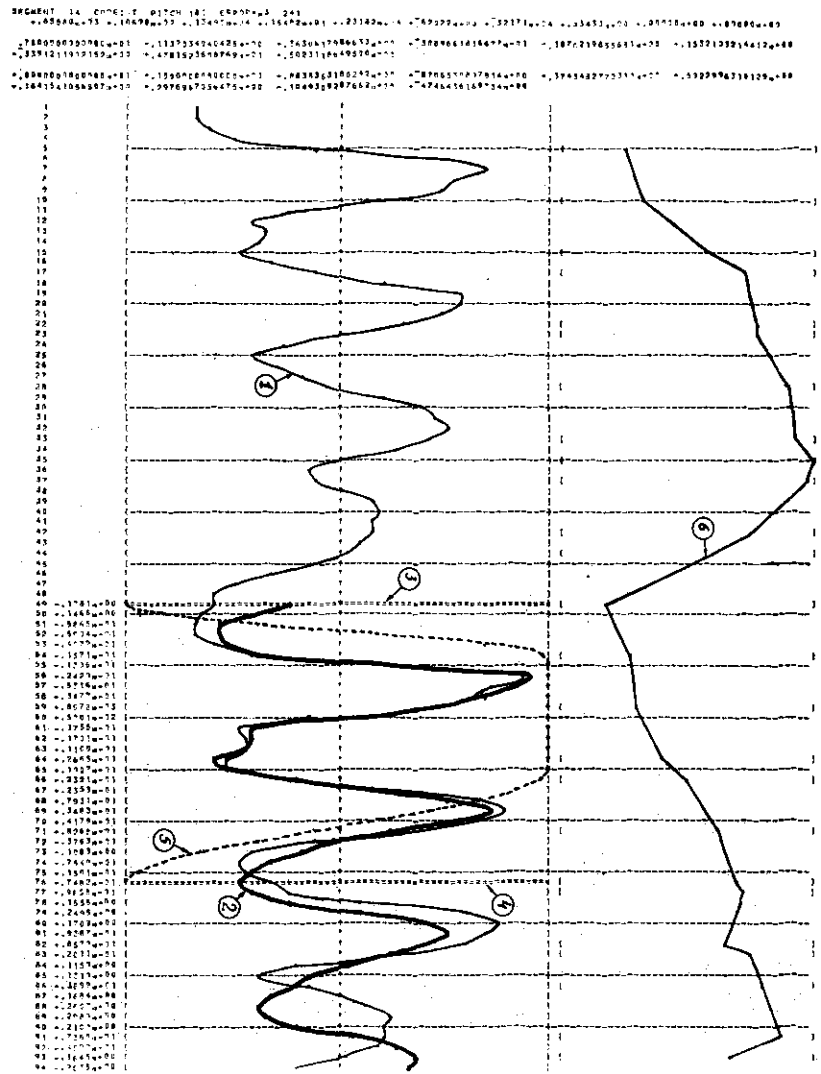


РИС. 1

положение максимума : $\beta 1_S$ = 0.2; значение в конце интервала : $\beta 3_S$ = 0.002.

Весовая функция при точной аппроксимации : Тип 4, $\beta 1$=0.2, $\beta 2$ = 0.6, $\beta 3$ = 0.006.

Кривая I-I*) - речевой сигнал после фильтрации до 3700гц с частотой квантования 8 кгц. Это "а" в слове "вада". Минимум кривой I-6, представляющей график взвешенной ошибки аппроксимации / экстраполяции, принят в качестве начала (I-3) оптимального интервала аппроксимации. I-5 - график взвешивающей функции при аппроксимации, I-2 - аппроксимирующий сигнал. Регуляризация существенного влияния на качество аппроксимации во временной области не оказывает. I-4 - конец интервала аппроксимации. Видно, что интервал закрытия голосовой щели заметно превышает интервал аппроксимации - открытие начинается в районе 85-89 отсчетов. Частота основного тона: I80 гц.

На рис. 2 фигурирует отрезок сигнала, представляющего конец второй фонемы "а" в том же слове "вада". Параметры анализа совпадают с приведенными выше, за исключением двух: порядок $D(z)$ как при поиске интервала, так и при самой аппроксимации принят равным четырем.

После завершения анализа печатается картинка видимой речи обработанного объекта, полученная с помощью программы "FACT-60" [I0]. Частоты каналов анализатора указаны в таблице I. С целью контроля для каждого сегмента, в котором была осуществлена аппроксимация, вычислялось положение полюсов передаточной функции. Эти полюса нумеровались в порядке возрастания частот, после чего были проведены линии, соединяющие соответствующие полюса для соседних сегментов. Следует подчеркнуть, что эта процедура выполнялась чисто интуитивно и, так же как и отпечатанные номера полюсов, не соответствует тем "истинным" формантам, которые определяются конфигурацией тракта диктора. Однако сходство полученных картинок и того, что принято считать формантным рисунком, особенно на рис. 4, очевидна. Число коэффициентов $D(z)$ при аппрокоимации было положено равным шести. На рис. 3 представлена нерегуляризованная картина; на рис. 4 - результаты анализа с параметром регуляризации 0.I. В нижней части рис. 3 и 4 расположены шесть колонок. Это :

- код типа сегмента (П, Т или Ш),
- номер сегмента (20 мсек - участка),

*) Первая цифра - номер рисунка, вторая - номер кривой.

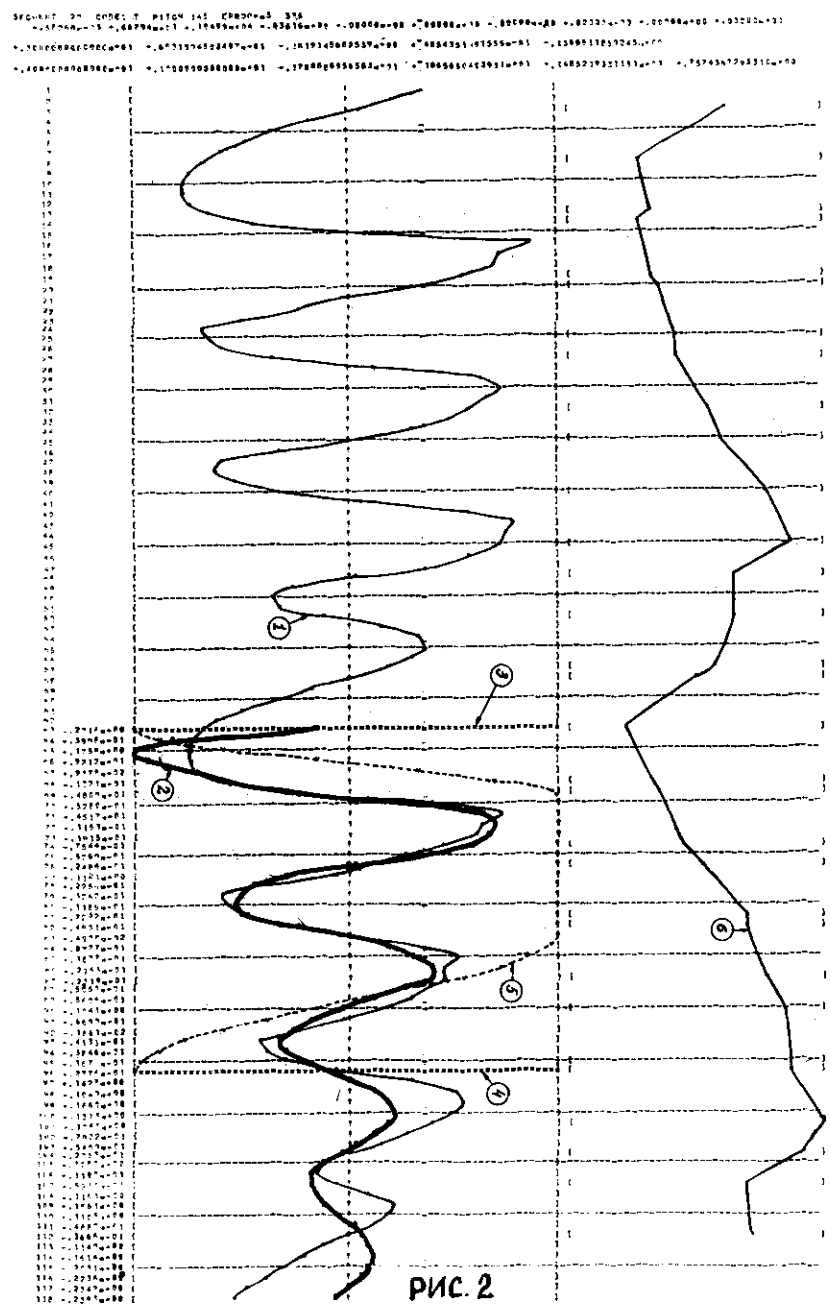


РИС. 2

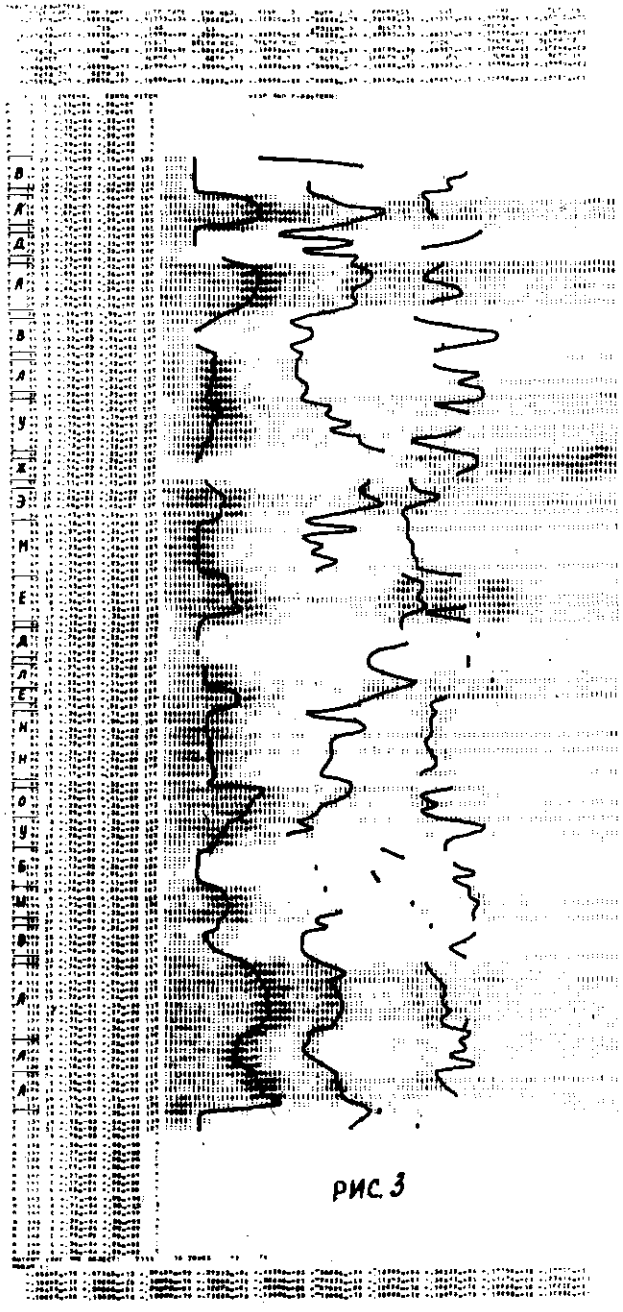


РИС. 3

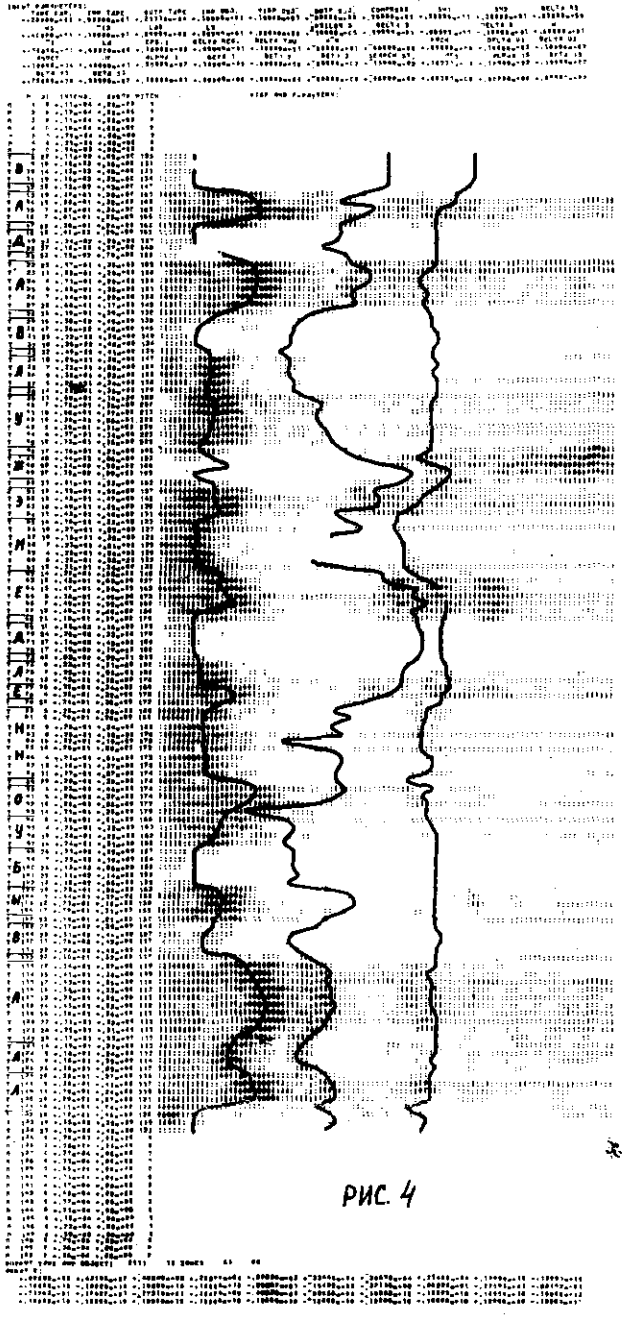


РИС. 4

Таблица I

Центральные частоты каналов (гц) на картинках "видимой речи"

номера каналов	0	I	2	3	4	5	6	7	8	9
0	-	75	98	I2I	I45	I70	I95	220	246	273
I0	300	328	356	385	4I5	445	476	507	540	572
20	606	640	675	7II	748	785	823	862	902	942
30	904	1026	1069	IIIЧ	II59	I205	I252	I300	I349	I399
40	I45I	I503	I556	I6II	I667	I724	I782	I84I	I902	I964
50	2027	2092	2I58	2225	2294	2364	2436	2509	2584	266I
60	2739	28I9	2900	2984	3069	3I56	3244	3335	3428	3522
70	3619	3717	38I8	392I	4026	4I33	4243	4355	4469	4586
80	4705	4827	4952	5079	5208	534I	5476	5615	5756	5900
90	6048	6I98	6352	6509	6669	6833	7000	-	-	-

Пример: центральная частота канала номер 63 : 2984 гц.

- индекс апериодичности, умноженный для удобства на сто,
- относительный уровень сигнала на сегментах,
- взвешенная ошибка аппроксимации,
- частота ОТ в герцах.

На рисунках хорошо заметно (сегменты II2-I32), что когда первые и вторые полюса близки, то последние располагаются выше спектральных максимумов, которые человек принял бы за проявление второй форманты. Это соответствует физике явления, поскольку для близко расположенных резонансных кривых положение максимумов смещается.

При синтезе речевого сигнала были выбраны следующие параметры : длительность импульса возбуждения $\tau_и = 0{,}2\ T_{от}$, форма импульса - треугольная, передний фронт $\tau_{пф} = 0{,}9\ \tau_и$.

Характеристики формирующего фильтра подбирались экспериментально - это нуль на частоте 0 гц шириной I0 гц, нуль на частоте 8 гц шириной I000 гц и полюс на частоте 3 кгц шириной I000 гц.

Форма сигнала возбуждения после формирующего фильтра приведена на графике 5 - I. 5-2 - отрезок синтезированного речевого сигнала. Это "а" в слове "убывала".

На качество синтезированной речи в большой степени влияет длительность импульса возбуждения и характеристики формирующего фильтра. Соотношение между фронтами возбуждающего импульса отражается на тембре речи, на осциллограмме, но незначительно влияет на качество.

Качество синтеза оценивалось путем неформального прослушивания и признано удовлетворительным. Случайное изменение в пределах нескольких процентов соседних периодов ОТ или длительностей импульса возбуждения не оказывало заметного влияния на качество. Особый интерес представляло сравнение качества синтеза по данным анализа с регуляризацией и без неё. Ощущается несколько большая плавность регуляризованной речи, но при этом ухудшается разборчивость, звук делается более глухим, машинноподобным.

Для корректной оценки объема информации, необходимого для передачи z-описания по линиям связи, необходимо было бы подвергнуть коэффициенты $D(z)$ одному из видов оптимального кодирования. На данном этапе проверено лишь усечение длины мантиссы при передаче коэффициентов b_j . Процедура эта выполня-

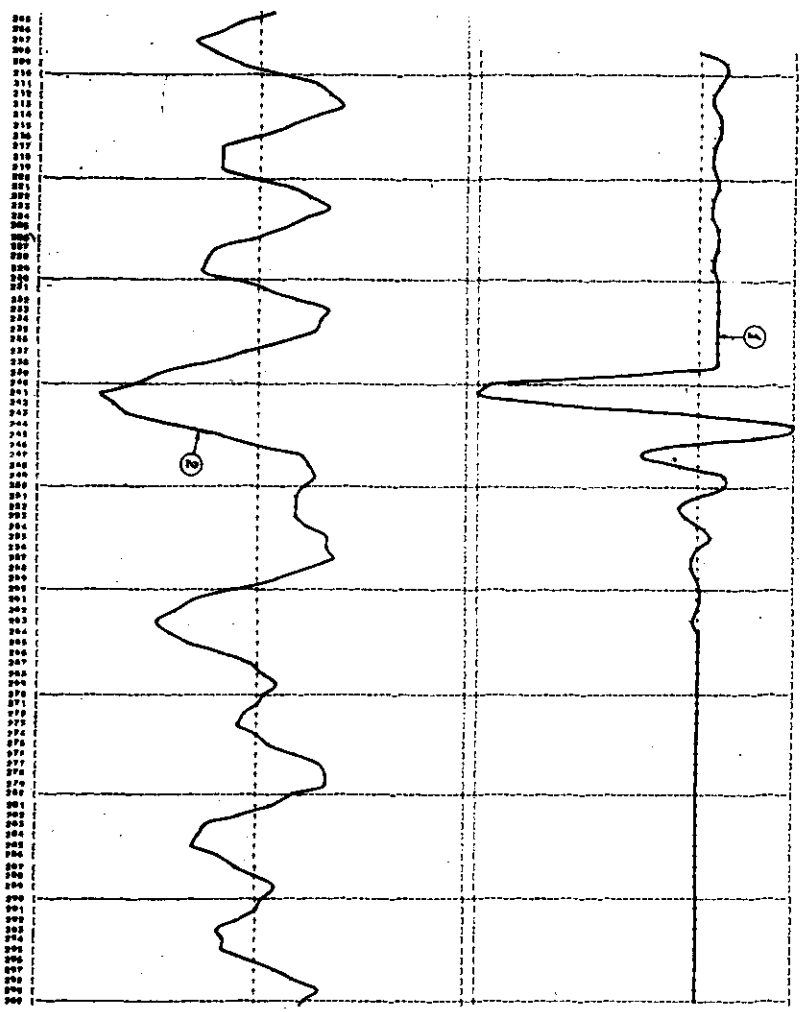


РИС. 5

лась следующим образом. Находился динамический диапазон изменения коэффициента на длительности фразы; значение каждого коэффициента для каждого сегмента с учетом найденных диапазонов приводилось к интервалу 0-.9999, переводилось в представление с фиксированной запятой, после чего сохранялось заданное в программе число старших разрядов мантиссы. Эксперимент показал, что при уменьшении числа бит на каждый коэффициент до I5 ухудшения качества синтезированной речи на слух не заметно. Небольшие искажения появляются при числе бит I0. При числе бит 8 и менее система не работает. Допустимое количество бит (I0) не зависит от степени используемой аппроксимации. Воспроизводился синтез при аппроксимации 8,6 и 4 порядка (соответственно 4,3 или 2 форманты). При $m = 4$ речь разборчива, но делается глуше. Возможно, в какой-то степени это поддастся корректировке при более тщательном подборе параметров источника возбуждения и формирующих фильтров.

Для более качественного звучания синтезированной речи следует ввести ещё одну кодировку типа сегмента : "К" - комбинированное (голос + шум) возбуждение. Это сделать, по-видимому, несложно, оценивая на стадии аппроксимации голосовых сегментов точность : если она ниже обычной, значит, в сигнале возбуждения присутствует шум.

6. Выводы

I. При работе на ЦВМ или специальной цифровой аппаратуре есть смысл пользоваться z - описанием дискретных систем, вместо того, чтобы допускать всевозможные неточности из-за попыток втиснуть в рамки дискретной техники непрерывные методы и описания.

Предлагается отказаться от попыток получить физиологически корректную формантную картину, учитывая трудности формализации процесса индексации спектральных максимумов. Предлагается синтезировать некую модельную (линейную динамическую) систему, добиться достаточного сходства сигналов во временной области, и z - параметры полученной передаточной функции модели использовать в качестве сокращенного описания речи.

2. Выяснено, что задача анализа, базирующаяся на среднеквадратической аппроксимации сигнала во временной области, является некорректной по Адамару с точки зрения вычисления поло-

жения полюсов (формант), т.е. близким сигналам во временной области могут соответствовать заметно различающиеся и в общем случае нестационарные формантные картинки. Отмечено, что для корректного в смысле реального физического речевого тракта формантного анализа недостаточно осциллограммы речи; кроме более подробных знаний об источнике, необходимо быть уверенным в наблюдаемости всех состояний системы. Этот путь для практических применений, например , в связи, неконструктивен.

3. Рассматриваются алгоритмы анализа речевого сигнала на основе z - описания [I] . Описан алгоритм поиска оптимального интервала аппроксимации, дающий хорошие результаты. Указан способ получения z -описания речи для сегментов с шумовым возбуждением.

4. Для контроля качества анализа была смоделирована на ЦВМ работа синтезатора. К достоинствам предлагаемого z - описания следует отнести очевидную простоту синтезирующей части. Получены первые экспериментальные результаты по синтезу.Как и ожидалось, регуляризация анализа, приводящая к сглаживанию формантных картинок, качества звучания не улучшила.Без дополнительных перекодировок при передаче z - параметров достаточно сохранять по I0 бит на каждый коэффициент. Рекомендуется использовать в описании передаточные функции шестой степени , однако , допустимо понижение степени до четвертой.

Литература

I. Лозовский В.С. Аппроксимация отклика системы в z -плоскости и формантный анализ речи. Сб. "Вычислительные системы", Новосибирск, № 37, I969.

2. Роберт Ли. Оптимальные оценки, определение характеристик и управление. Москва, Физматгиз "Наука", I966.

3. G.Golub, Numerical Methods for Solving Linear Least Squares Problems, Numerische Mathematik, B.7, H.3, 1965.

4. Тихонов А.Н. О некорректных задачах линейной алгебры и устойчивом методе их решения. ДАН СССР, т.I63, № 3, I96[illegible].

5. Лозовский В.С. Модифицированный разностный метод определения основного тона речи. "Труды акустического института", Москва, вып. ХП, I970.

6. Лозовский В.С. Программа синтеза рекурсивных цифровых фильтров. Сб. "Вычислительные системы", Новосибирск, № 35, I969.

7. Лозовский В.С. Метод получения полюсно-нулевого описания речевого сигнала (Аннотация доклада на Всесоюзной школе-семинаре АРСО-5, Сухуми, I969 г.), Труды Акустического института, М., вып. ХП, I970.

8. Лозовский В.С. Процедура решения полиномиальных уравнений методом Мюллера. (наст. сборник).

9. Фант Г. Анализ и синтез речи, пер. с англ. под редакцией Н.Г. Загоруйко, Новосибирск, "Наука", I970.

I0. Лозовский В.С. Программа спектрального анализа в частотной и временной областях. Сб. "Вычислительные системы", Новосибирск, I97I, (в печати).

Поступила в редакцию
I3.I.I97I г.