МЕТОДЫ ОБНАРУЖЕНИЯ ЗАКОНОМЕРНОСТЕЙ С ПОМОЩЬЮ ЭВМ
(Вычислительные системы)
1981 год Выпуск 91

УДК 519.226:534.44

АЛГОРИТМ ВЫДЕЛЕНИЯ ОСНОВНОГО ТОНА ПО РАЗНОСТНОЙ ФУНКЦИИ РЯДА ОСТАТОЧНЫХ ОШИБОК МОДЕЛИ АВТОРЕГРЕССИИ

А.В.Кельманов

I. Постановка задачи

Как известно [I], речь образуется (в первом приближении) в результате свертки функции возбуждения с импульсной реакцией речевого тракта. Для голосовых участков речи функция возбуждения предс: авляет собой случайный импульсный процесс с детерминированным тактовым интервалом, который называется периодом основного тона (обратная величина называется частотой основного тона). Основную фонетическую информацию в речевом сигнале несет импульсная реакция, тогда как в функции возбуждения заключена лишь частичная информация о фонетических признаках. Однако функция возбуждения содержит в себе информацию об индивидуальных особенностях голоса говорящего [2]. Например, каждому человеку присуща характерная интонационная окраска его голоса, которая формируется именно функцией возбуждения и определяется изменением частоты основного тона во времени. Кроме того, в траектории частоты основного тона содержатся данные об эмоциональном состоянии говорящего [3]. В свою очередь, информация об индивидуальных особенностях голоса и эмоциональном состоянии говорящего учитывается в том или ином виде почти во всех человеко-машинных системах, а в некоторых речевых приложениях представляет самостоятельный интерес. Поэтому выделение частоты основного тона является важной задачей.

Целью настоящей работы является разработка алгоритма оценивания частоты или периода основного тона по акустическому речевому сигналу, представленному в цифровом виде.

Будем считать, что передаточная функция речевого тракта содержит только полюсы:

$$H_{p_T}(z^{-1}) = \frac{1}{E(z^{-1})} = \frac{1}{1 - \sum_{i=1}^{p_T} e_i z^{-i}}, \qquad (1)$$

а импульсная реакция речевого тракта h_n описывается уравнением авторегрессии:

$$h_n = \sum_{i=1}^{p_T} e_i h_{n-i} + \varepsilon_n, \qquad (2)$$

где ε_n - ошибка, так что уравнение речеобразования в терминах z-преобразования может быть записано в виде

$$H^{-1}_{p_T}(z^{-1}) s_n = E(z^{-1}) s_n = y_n, \qquad (3)$$

где s_n - отсчеты речевого сигнала, y_n - отсчеты функции возбуждения, а n - номер отсчета. Обозначим через T_0 период основного тона. Тогда момент возникновения m-го импульса $t_m = mT_0 + \nu$, где ν - случайная величина с матожиданием $M[\nu] = 0$, $|\nu| = T_0/c$, $c \geq 2$, а $T_{min} \leq t_m - t_{m-1} \leq T_{max}$. Далее будем считать, что форма импульсов возбуждения известна приближенно и на интервалах анализа длительностью $T_a = \lambda T_0$, где $2 \leq \lambda \leq 4$, изменяется достаточно мало, так что $y_n \approx y_{n-k}$, где k - число отсчетов в период основного тона.

Итак, пусть спектр речевого сигнала s_n ограничен частотой $F_s/2$, а сам сигнал квантован с частотой F_s и задан в виде временного ряда $\{s_n\}$. Требуется на интервалах анализа (сегментах) $\{s_n\}^l$, $n = \overline{1,N}$, $l = \overline{1,L}$, полученных путем разбиения сигнала на L отрезков длительностью T_a и перекрывающихся на время $\Delta T < T_a$, определить параметр k или период основного тона T_0, или частоту основного тона F_0, если интервалы относятся к классу голосовых.

2. Методы выделения основного тона

История проблемы выделения основного тона насчитывает несколько десятков лет. К настоящему времени разработано так много методов выделения основного тона, что даже краткая их характеристика потребовала бы написания не одной обзорной статьи. К счастью, исчерпывающий обзор по данной проблеме к началу 70-х годов сделан в [4]. Казалось бы, из большого числа имеющихся методов можно найти подходящий. Однако, как показывает динамика публикаций последних лет, интерес к методам выделения основного тона не ослабевает

и по сей день: идет поиск новых и усовершенствование старых методов. При этом основные усилия направлены на повышение надежности и сокращение трудоемкости. К наиболее часто используемым и хорошо себя зарекомендовавшим относятся следующие методы: 1) автокорреляционные [4,5,13], 2) сдвиговые или разностные [6,7], 3) кепстральные [12], 4) авторегрессионные [8,14,16].

Качество методов выделения основного тона зависит от того, насколько хорошо из речевого сигнала удалены характеристики речевого тракта ($H_{рт}(z^{-1})$), выполняющего преобразование (в данном случае шумящее) сигнала голосовых связок. Поэтому автокорреляционные и сдвиговые методы выделения основного тона непосредственно по речевому сигналу s_n, не устраняющие влияние речевого тракта, менее надежны по сравнению с методами, позволяющими произвести хотя бы частичное его устранение. Так, например, использование клипирования речевого сигнала позволило повысить надежность автокорреляционных и сдвиговых методов [7,13]. Хотя кепстральные методы и являются самыми трудоемкими, но до последнего времени они применяются наиболее часто именно потому, что позволяют разделить характеристики голосового источника и речевого тракта с устранением последних и, таким образом, позволяют получить высокую надежность оценивания. В настоящее время наиболее перспективными считаются авторегрессионные методы, так как они, по сравнению с кепстральными, при весьма незначительном понижении надежности требуют существенно меньших затрат времени [15]. Если учесть, что авторегрессионные методы еще до конца не исследованы, то становится понятным повышенный интерес к ним.

Идея авторегрессионных методов состоит в следующем. Речевой сигнал s_n аппроксимируется моделью авторегрессии p-го порядка

$$A(z^{-1})s_n = \alpha_n, \quad n = \overline{1,N}, \qquad (4)$$

где $A(z^{-1}) = 1 - \sum_{i=1}^{p} a_i z^{-i}$ - оператор авторегрессии, α_n - ошибка аппроксимации, z^{-1} - оператор сдвига назад такой, что $z^{-1}s_n = s_{n-1}$.

Если при аппроксимации порядок модели (4) p выбрать больше порядка $p_т$ оператора $E(z^{-1})$ в (1), то из (1)-(4) следует, что функция возбуждения y_n вынужденно аппроксимируется моделью авторегрессии $(p - p_т)$-го порядка с оператором $E_0(z^{-1})$, т.е. предполагается, что $E(z^{-1})E_0(z^{-1}) = A(z^{-1})$. Поэтому, во-первых, в ряде

остаточных ошибок $\alpha_n = \hat{A}(z^{-1})s_n$ исключаются характеристики речевого тракта, а во-вторых, из-за того, что $p \ll k$ и $N = \lambda k$, ряд остаточных ошибок $\{\hat{\alpha}_n\}$ будет обладать свойством периодичности. Эти свойства ряда $\{\hat{\alpha}_n\}$ позволяют эффективно использовать идеи авто - корреляционных и сдвиговых методов. Так, в алгоритме SIFT [I4] основной тон определяется по автокорреляциям ряда остаточных ошибок $\{\hat{\alpha}_n\}$, а в алгоритме LPCAMDF [I6] - по сдвиговой функции того же ряда.

Почти все алгоритмы выделения основного тона предполагают низкочастотное ограничение спектра речевого сигнала, как правило, частотой I кгц, и понижение исходной частоты квантования F_s в соответствии с теоремой Котельникова до 2 кгц. Это делается с целью сокращения объема вычислений и частичного устранения характерис - тик речевого тракта путем фильтрации высокочастотных составляющих сигнала. Последнее объясняется тем, что основная информация о голосовом источнике заключена в низкочастотной области, а составляющие верхних частот при выделении основного тона являются шумящими и, как правило, вносят ошибки. От частоты квантования сигнала зависит, с какой погрешностью будет выделен основной тон. Поскольку максимальная погрешность в измерении частоты основного тона равна $1/2F_sT_0^2$, понижение частоты квантования приводит к уменьшению разрешающей способности частотной шкалы, по которой производится измерение. Нетрудно заметить, что при F_s = 2 кгц и T_0 = 2-I5 мсек погрешность лежит в пределах I,I-62,5 гц, т.е. может достигнуть значительной величины. Следовательно, низкочастотное ограничение спектра сигнала и понижение частоты квантования уменьшает вероятность ошибки, вносимой высокочастотными компонентами, и сокращает трудоемкость, но в то же время увеличивает погрешность оценивания.

Один из способов уменьшения суммарной ошибки определения основного тона состоит в уменьшении погрешности оценивания или в увеличении разрешающей способности шкалы измерений путем интерполяции. В алгоритме SIFT, например, для этой цели используется тригонометрическая интерполяция. Другой способ состоит в том, что по сравнению с предыдущим вместо интерполяции используется менее сильное ограничение спектра сигнала и более высокая частота квантования, позволяющая повысить разрешающую способность. Этот способ использован в алгоритме LPCAMDF. В этом алгоритме спектр сигнала ограничивался частотой 3,2 кгц, а частота квантования сигнала равнялась 6,8 кгц. Поэтому при T_0 = 2-I5 мсек погрешность из-

мерения лежит в диапазоне 0,3-18,4 гц. Такие значения погрешности в некоторых приложениях являются недопустимыми, поэтому и для данного способа необходима интерполяция. К тому же указанный способ из-за наличия в сигнале высокочастотных компонент, вносящих ошибки, требует использования сложного блока коррекции, сглаживания и исправления ошибок и, как показывает практика, является менее перспективным.

Одним из недостатков большинства алгоритмов выделения основного тона, в том числе и алгоритмов SIFT и LPCAMDF, является то, что они при неплохой надежности оценивания на голосовых участках весьма ненадежно разделяют голосовые и неголосовые звуки и поэтому в конечном итоге могут привести к значительным ошибкам оценивания траектории основного тона. Кроме того, в этих алгоритмах вычисление периода фактически производится не только на голосовых участках, где это требуется, но и на шумовых, где оно не нужно, т.е. тратится лишнее время.

В данной работе предлагается алгоритм, лишенный указанных недостатков, он является модификацией алгоритмов SIFT, LPCAMDF и алгоритма, описанного в [7].

3. Алгоритм

Итак, пусть речевой сигнал разбит на L сегментов $\{ s_n \}^l$, $n = \overline{1,N}$, $l = \overline{1,L}$.

Оценивание основного тона производится последовательно, начиная с сегмента с номером I и кончая сегментом с номером L. Соотношения, выписанные ниже, приведены для произвольного сегмента с номером 1. Индекс 1 в них для простоты опущен. Но в соотношениях, где результат операции, выполняемой для 1-го сегмента, зависит от результатов для предшествующих сегментов, индексы этих сегментов указаны.

На первом шаге алгоритма производится низкочастотная фильтрация при помощи чебышевского фильтра 3-го порядка с частотой среза 0,8 кгц, а исходная частота квантования F_s понижается до 2 кгц так, что объем выборки N сигнала $\{s_n\}$ для профильтрованного сигнала $\{x_n\}$ сокращается до величины

$$N_1 = (N-1)/k_v + 1, \quad k_v = F_s/2. \qquad (5)$$

Далее производится классификация ряда $\{x_n\}$ при помощи алгоритма тон/шум, описанного в [10]. Напомним, что решение тон/шум в

этом алгоритме принимается по критерию χ^2, а статистика построена по частным автокорреляциям модели авторегрессии. При этом переменной IV присваиваются значения

$$IV = \begin{cases} 1, \text{ если сегмент голосовой,} \\ 0 - \text{в противном случае.} \end{cases} \qquad (6)$$

Если IV = 0, то полагаем равными нулю число отсчетов в периоде основного тона k и значения периода T_0 и частоты F_0 основного тона. Если IV = 1, то производится оценивание параметра k. Для этого сигнал $\{x_n\}$ аппроксимируется моделью авторегрессии со стационарной разностью первого порядка:

$$A(z^{-1})w_n = \alpha_n, \qquad w_n = \nabla x_n, \quad n = \overline{1, N_1}. \qquad (7)$$

Здесь $\nabla = 1 - z^{-1}$ – оператор взятия разности такой, что $\nabla x_n = x_n - x_{n-1}$. Операция взятия разности вводится с целью частичного устранения влияния голосового источника [11] на оценивание параметров речевого тракта с тем, чтобы в проинтегрированном ряде остаточных ошибок

$$\hat{u}_n = \nabla^{-1}\hat{\alpha}_n = x_n - \sum_{i=1}^{p} \hat{a}_i x_{n-i}, \quad n = \overline{1, N}, \qquad (8)$$

наиболее полно устранить характеристики речевого тракта. Оценки параметров $\{\hat{a}_i\}$ находятся из решения нормальных уравнений методом Дурбина [9].

После вычисления ряда $\{\hat{u}_n\}$ производится вычисление разностной функции

$$d(i-n_1+1) = \sum_{j=1}^{n_{12}} |\hat{u}(j) - \hat{u}(j+i)|, \quad i = \overline{n_1, n_2};$$

$$n_{12} = n_2 - n_1 + 1; \; n_1 = (N_{min} - 1)/k_v + 1; \; n_2 = (N_{max} - 1)/k_v + 1; \qquad (9)$$

$$N_{min} = T_{min}F_s + 1; \; N_{max} = T_{max}F_s + 1,$$

которая является мерой сходства между участками сигнала $\hat{u}_n$. При чисто периодическом сигнале $\hat{u}_n$ функция (9) будет иметь четко выраженные минимумы, в которых ее значение равно нулю. Чем сильнее сигнал отличен от периодического, тем слабее выражены минимумы и тем больше значение функции (9) в точках минимума.

Поиск минимума и абсциссы минимума осуществляются по параболически скорректированной разностной функции

$$d_0(j) = d(j) + (j-\beta^*)^2 \cdot \gamma^* + j\cdot\rho, \quad j = \overline{1, n_{12}}, \tag{10}$$

т.е.

$$d_{0min} = \min_{j \overline{n_1, n_2}} d_0(j); \quad \beta = \operatorname{argmin} d_0(j). \tag{11}$$

Здесь $\beta^*=\beta(1-1)$ – абсцисса минимума для предыдущего сегмента, $\rho > 0$ – коэффициент перекоса, $\gamma^* = \gamma(1-1)$, а γ – коэффициент параболической коррекции, определяющийся из соотношения:

$$\gamma = \begin{cases} \gamma(1-1) + [\gamma_{max} - \gamma(1-1)]\cdot\phi_1, & \text{если } IV(1-1) = 1, \\ \gamma(1-1) + [\gamma_{min} - \gamma(1-1)]\cdot\phi_2, & \text{если } IV(1-1) = 0, \end{cases} \tag{12}$$

где γ_{max} и γ_{min} – желаемые пределы изменения коэффициента γ, а ϕ_1 и ϕ_2 – коэффициенты, определяющие скорость возрастания и убывания γ. Пределы m_1 и m_2 поиска минимума вычисляются адаптивно, в зависимости от числа отсчетов k в периоде основного тона, найденного на предыдущем сегменте:

$$\begin{aligned} m_1 &= \max\{1, k(1-1)/k_v - m_a - n_1 + 1\}, \\ m_2 &= \min\{n_{12}, k(1-1)/k_v - m_a - n_1 + 1\}. \end{aligned} \tag{13}$$

Параметр m_a задает интервал поиска минимума.

После того как найдена абсцисса минимума β функции $d_0(j)$, может быть вычислено число отсчетов в периоде профильтрованного сигнала $\{x_n\}$. С учетом (9) это число равно $\beta+n_1-1$.

Далее производится параболическая интерполяция по трем точкам $d_0(\beta-1), d_0(\beta)$ и $d_0(\beta+1)$ с шагом $1/k_v$ с целью определения абсциссы "истинного" минимума, и, таким образом, разрешающая способность измерения увеличивается до разрешающей способности, задаваемой исходной частотой квантования. Значение параметра k находится из соотношения:

$$k = (\beta+n_1-1)\cdot k_v + k_n, \tag{14}$$

в котором через k_n обозначена поправка, вычисленная на этапе интерполяции. Затем осуществляется коррекция ошибок по правилам:

$$\forall l\{l > 2 \,\&\, k(l-1) = 0 \,\&\, k(l-2) = 0 \Rightarrow k(l-1) = [k(l-2) + k(l)]/2\},$$
$$\forall l\{l > 2 \,\&\, k(l-1) \neq 0 \,\&\, k(l) - k(l-1) > k_{\nabla} \Rightarrow k(l) = k(l-1)\}. \quad (15)$$

Поиск основного тона осуществляется при следующих начальных условиях: $m_1(1) = 1$; $m_2(1) = n_{12}$; $\beta^* = \beta(0) = (n_2 - n_1)/2$; $\gamma^* = \gamma(0) = \gamma_{min}$. Наконец, период и частота основного тона вычисляются по соотношениям:

$$T_0 = (k-1)/F_s; \quad F_0 = 1/T_0.$$

Таким образом, работа алгоритма заключается в последовательном выполнении соотношений (5)-(I6).

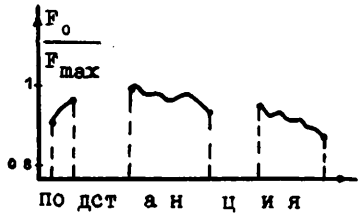
4. Эксперименты и их результаты

Р е ч е в о й м а т е р и а л. При проведении эксперимента использовался словарь, состоящий из 45 слов. По этому словарю тремя дикторами-мужчинами было сформировано 6 акустических последовательностей (каждый диктор произносил слова дважды). Общее число участков анализа - около I5 тыс. Кроме того, тестирование проводилось на искусственном голосовом речеподобном сигнале. Синтез осуществлялся так же,как и в работе [II].

У с л о в и я э к с п е р и м е н т а. Описанный алгоритм, а также алгоритмы SIFT и LPCAMDE были запрограммированы на языке "Фортран" для ЭВМ "Минск-32". Ввод слов в ЭВМ осуществлялся в машинном зале (уровень шума ~60 дб) через микрофон типа МД-59 и семиразрядный аналого-цифровой преобразователь при частоте кванто - вания 20 кгц.

У с л о в и я а н а л и з а. Акустические реализации естественного и искусственного сигналов были обработаны при помощи выше описанной процедуры, а также алгоритмов SIFT и LPCAMDE . Обработка сигналов проводилась при длительности сегмента T_a=25,6 мсек и сдвиге от сегмента к сегменту на время ΔT = I6 мсек.Каждый сегмент перед обработкой взвешивался окном Хэмминга. Значения параметров, управляющих работой алгоритма, были следующими: n_1=5; n_2= = 25; $\gamma_{min} = 7{,}5 \cdot 10^{-5}$; $\gamma_{max} = 1{,}25 \cdot 10^{-3}$; $\phi_1 = 0{,}3$; $\phi_2 = 0{,}2$; $\rho = 0{,}006$; $p = 4$; $m_a = 4$.

Р е з у л ь т а т ы о ц е н и в а н и я. Для примера на рисунке изображена траектория изменения частоты основного тона



нормирования к I по максимальному значению для слова "подстанция" (диктор I), полученная при помощи описанного алгоритма. Результаты алгоритмического оценивания сравнивались с результатами визуального оценивания по осциллограмме акустической реализации. При этом для описанного алгоритма расхождений почти не наблюдалось, в то время как для алгоритмов SIFT и LPCAMDF были замечены значительные расхождения. Проверка всех алгоритмов на синтезированном голосовом речеподобном сигнале показала их практическую равноценность. Поэтому более тщательно были проверены блоки тон/шум. Оказалось, что в алгоритмах SIFT и LPCAMDF эти блоки дают ошибки примерно в 5,7% случаев, а в предлагаемом алгоритме только в I,3% случаев.

5. Обсуждение результатов

Разработка алгоритма, описанного в данной работе, была обусловлена тем, что, во-первых, описание речевого сигнала моделью авторегрессии со стационарной сезонной разностью, в которой число отсчетов в периоде основного тона является одним из параметров, позволило повысить надежность оценивания параметров речевого тракта [II]. Во-вторых, оказалось, что если при аппроксимации речевого сигнала моделью авторегрессии с сезонной разностью воспользоваться алгоритмом SIFT, то параметры речевого тракта, получающиеся в результате оценивания, в ряде случаев ведут себя весьма нестабильно во времени [II], что не согласуется с физиологическими данными [I]. Следовательно, причина нестабильности кроется в алгоритме SIFT (при устойчивом методе оценивания). В-третьих, проведенные "усовершенствования" алгоритма SIFT не оправдали себя: надежность оценивания не повысилась. В-четвертых, априори было неясно, с какой степенью точности необходимо выделять основной тон для анализа и распознавания речи при помощи модели авторегрессии с сезонной разностью, чтобы, например, при распознавании сделать выводы о самой модели, исключив ошибки метода выделения основного тона. Наконец, кепстральный алгоритм был отклонен из-за его высокой трудоемкости, а алгоритм, описанный в [7], и алгоритм LPCAMDF были отклонены по той причине, что при пробной проверке на тестовых акустических реализациях они приводили к бо́льшим, чем алгоритм SIFT, ошибкам.

Ранее было отмечено, что предложенный в настоящей работе алгоритм является модификацией трех алгоритмов. В нем, как и в алгоритмах SIFT и LPCAMDF, оценивание основного тона производится по ряду остаточных ошибок и, как в алгоритме LPCAMDF и в алгоритме, описанном в [7], в качестве меры "сходства" участков речевого сигнала служит сдвиговая, или разностная функция. Основные отличия предлагаемого алгоритма от трех упомянутых состоят в том, что в нем разделены оценивание и классификатор тон/шум, а сам классификатор построен по новому принципу. Кроме того, в описанном алгоритме применена более простая в вычислительном плане параболическая интерполяция и, наконец, сигнал аппроксимировался моделью авторегрессии в проинтегрированной форме.

Применение модели авторегрессии в проинтегрированной форме вместо модели авторегрессии позволяет повысить надежность оценивания основного тона за счет сглаживания шумящих высокочастотных компонент. Далее заметим, что в алгоритме, предложенном в [7], и алгоритме LPCAMDF решение тон/шум производится путем сравнения с порогом нормированной разностной функции и энергии сигнала на сегменте. Энергия сигнала на сегменте не позволяет надежно классифицировать звуки на звонкие и глухие, так как на практике уровни энергий для типов звуков довольно часто совпадают из-за различной громкости произнесения. По этой же причине затруднено принятие решения по ненормированной разностной функции. Несмотря на ее нормирование, проводящееся в алгоритме, предложенном в [7], и алгоритме LPCAMDF, значение порога при классификации, как показывают результаты экспериментальных исследований, приходится перестраивать, хотя и незначительно, не только от диктора к диктору, но даже от одной акустической реализации к другой для одного и того же диктора. При этих же условиях классификатор описанного алгоритма менее критичен, а разделение классификации и оценивания позволяет обойтись без нормализации разностной функции. Для значений периода основного тона, лежащих в диапазоне 2-15 мсек, разрешающая способность описанного алгоритма равна 0,1-6,3 гц.

Трудоемкость алгоритма складывается из трудоемкостей классификации тон/шум и оценивания. Для проведения классификации необходимо $N_1(K+1) + K^2 + K - p + 5$ операций, где K - число частных автокорреляций, по которым принимается решение [10]. Для оценивания периода без учета интерполяции и коррекции необходимо: N_1 операций на взятие разностей, $2N_1(p+1) + p^2$ операций на вычисление автокорреляций, параметров авторегрессии и ряда остаточных ошибок,

$4n_{12}$ операций на вычисление разностной функции и ее параболическую коррекцию и $2m_a$ операций на поиск минимума. Поэтому трудоемкость алгоритма равна $N_1(2p+K+4) + K^2 + p^2 + K + 4n_{12} + 2m_a - p + 5$ операций сложения и умножения.

Трудоемкости алгоритмов SIFT и LPCAMDF без учета интерполяции и коррекции практически одинаковы, равны $N_1(2p + n_{12} + 3) + p^2 + 2m_a + 5$ операций и при реальных значениях параметров меньше, чем трудоемкость описанного алгоритма. Однако разность между трудоемкостями незначительна.

Остается отметить, что в настоящее время алгоритм успешно используется для анализа и распознавания речи. В качестве примера приведем тот факт, что введение нормированной частоты основного тона в качестве признака в систему распознавания изолированных слов, работающую без подстройки под диктора, позволило повысить надежность распознавания примерно на 4%.

Таким образом, предложенный в настоящей работе алгоритм по сравнению с известными аналогами позволяет при незначительном увеличении трудоемкости повысить надежность оценивания периода основного тона за счет уменьшения примерно в 4,4 раза числа ошибок классификации тон/шум.

Литература

1. САПОЖКОВ М.А. Речевой сигнал в кибернетике и связи. - М.: Связьиздат, 1963. - 450 с.

2. РАМИШВИЛИ Г.С. Речевой сигнал и индивидуальность голоса. -Тбилиси: Мецниереба, 1976. - 183 с.

3. Речь, эмоции и личность: Материалы Всесоюзного симпозиума, 27-28 февраля 1978 г. - Л.: 1978. - 186 с.

4. Вокодерная телефония /Под ред. Пирогова А.А. - М.: Связь, 1974. - 535 с.

5. БАРОНИН С.П. Автокорреляционный метод выделения основного тона речи. -В кн.: Тр. Гос.НИИ связи СССР, М., 1961, вып. 3, с.93-102.

6. СОБОЛЕВ В.Н., БАРОНИН С.П. Исследование сдвигового метода выделения основного тона речи. - Электросвязь, 1968, № 12, с.33-36.

7. ЛОЗОВСКИЙ В.С. Модифицированный разностный метод определения основного тона речи. - Тр. Акустического института, 1974, вып. XII, с.189-193.

8. АКИНЬИЕВ Н.Н. Устройство для выделения артикуляционных сигнал-параметров и сигнал-остатка речевого сигнала. Авт.св. СССР № 142698. - Бюл. Открытия. Изобретения. Промышленные образцы. Товарные знаки. 1961, № 22, с.37.

9. БОКС Дж., ДЖЕНКИНСК Г. Анализ временных рядов, прогноз и управление. Т.I. - М.: Мир, 1974. - 406 с.

I0. КЕЛЬМАНОВ А.В. Алгоритм классификации тон/шум по частным автокорреляциям. -В кн.: Эмпирическое предсказание и распознавание образов (Вычислительные системы, вып. 83). Новосибирск, I980, с. 67-73.

II. КЕЛЬМАНОВ А.В. Оценивание параметров речевого тракта в классе моделей авторегрессии со стационарной сезонной разностью первого порядка и стационарной разностью первого порядка.-В кн.: Эмпирическое предсказание и распознавание образов. (Вычислительные системы, вып. 76). Новосибирск, I978, с.II0-I3I.

I2. NOLL A.M. Cepstrum pitch determination.- The Journal of the Acoustical Society of America,1967,v.41,N 2,p.293-309.

13. SONDHI M.M. New methods of pitch extraction.- IEEE Trans. Audio Electroacoust.,1968,v.AU-16,N 3,p.262-266.

14. MARKEL J.D. The SIFT algorithm for fundamental Fregnency estimation. - IEEE Trans.Audio Electroacoust.,1972,v.AU-20, N 5, p.367-377.

15. A Comparative performance study of several pitch detec - tion algorithms/ Rabiner L.R., Cheng M.J., Rosenberg A.E., McGonegal C.A. - IEEE Trans.Acoust., Speech and Signal Processing,1976, v.ASSP-24,N 5,p.399-418.

16. UN C.K.,YANG S.C. A pitch extraction algorithm based on LPC inverse filtering and AMDF. - IEEE Trans.Acoust., Speech and Signal Processing,1977,v.ASSP-25,N 6,p.565-572.

Поступила в ред.изд.отд.
I7 ноября I979 года