

ОБРАБОТКА АКУСТИЧЕСКИХ СИГНАЛОВ. КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ

УДК 612.85

ОЦЕНКА ТРЕКОВ РЕЗОНАНСНЫХ ЧАСТОТ РЕЧЕВОГО ТРАКТА

© 2023 г. А. С. Леонов^а, В. Н. Сорокин^{б, *}

^аНациональный исследовательский ядерный университет “МИФИ”,
Каширское ш. 31, Москва, 115409 Россия

^бИнститут проблем передачи информации им. А.А. Харкевича РАН,
Большой Каретный пер. 19, стр. 1, Москва, 127051 Россия

*e-mail: vns@iitp.ru

Поступила в редакцию 23.01.2023 г.

После доработки 23.01.2023 г.

Принята к публикации 18.05.2023 г.

Предлагается новый метод оценки треков формантных частот речевого тракта для произвольных речевых сегментов. Метод использует отношение двух преобразований Фурье речевого сигнала со специальными окнами экспоненциального типа, зависящими от некоторого параметра. Это отношение используется для определенных моментов времени и рассматривается как функция частоты и параметра. Анализируя для нескольких значений параметра распределение точек минимума (по частоте) для фазы этого отношения и/или аналогичное распределение точек экстремума для его амплитуды, можно оценить формантные частоты по пикам этих распределений. Представлено математическое исследование, обосновывающее такой подход. Проведены серии численных экспериментов по обработке синтетических и реальных речевых сигналов, подтвердившие работоспособность предложенного метода оценки формант. В частности, в экспериментах с синтезированными гласными было установлено, что погрешность оценки их резонансных частот мала и устойчива по отношению к аддитивным шумам вплоть до отношения сигнал/шум +5 дБ. Для реальной речи метод позволяет вычислить треки формантных частот как для звуков с голосовым возбуждением, так и для глухих фрикативных, аспиративных взрывов и шепотной речи.

Ключевые слова: анализ речевого сигнала, формантные частоты, экстремумы действительной и мнимой компонент спектра

DOI: 10.31857/S0320791923600749, **EDN:** CPWOFD

1. ВВЕДЕНИЕ

Резонансные частоты речевого тракта содержат информацию, необходимую для распознавания фонетических элементов речи, определения диалекта и акцента носителя иностранного языка. Распознавание возраста, пола и личности диктора, его эмоционального и психофизического состояния также в значительной степени опирается на анализ резонансных частот. Роль резонансных частот при распознавании пола диктора особенно велика, если частота основного тона его речи лежит в области пересечения распределений частот основного тона голосов мужчин и женщин.

Резонансные частоты определяются длиной и формой речевого тракта, положением небной занавески, площадью голосовой щели и податливостью стенок тракта. На динамику резонансных частот влияют механические параметры артикуляторных органов (масса и упругость присоединенных тканей), свойства управляющих мышц и характеристики системы управления артикуляцией. Поэтому резонансные частоты служат ос-

новным источником информации при решении обратных задач определения формы речевого тракта, положений артикуляторов и команд управления артикуляторами.

На звуках с голосовым возбуждением эти частоты целесообразно искать синхронно с периодом основного тона на интервалах закрытой голосовой щели. Резонансы проявляются в виде пиков в амплитудно-частотном спектре речевого сегмента — формант, частота которых близка к резонансным частотам речевого тракта.

Форманты наблюдаются не только в спектрах звуков с голосовым возбуждением, но и в спектрах глухих фрикативных, а также в спектрах так называемого взрыва — участка речевого сигнала после раскрытия смычки взрывных звуков /*n, m, k, b, d, g*/. Частота формант взрыва в некоторой степени определяет место артикуляции (положение смычки в речевом тракте) взрывных звуков [1]. Но в слитном потоке речи этой информации недостаточно для распознавания таких звуков. Необходимо использовать еще и резонансные ча-

стоты непосредственно перед смычкой, а также учитывать их динамику до и после смычки. Наиболее надежно место артикуляции взрывных находится путем решения обратной задачи относительно площади поперечного сечения речевого тракта по резонансным частотам [2, 3].

Вследствие важной роли резонансных частот в процессах речеобразования в речевых исследованиях было разработано множество методов формантного анализа речевых сигналов, прежде всего, в спектрально-временной области. Обзор и подробный анализ методов определения резонансных частот в спектрально-временной области даны в [3]. Несмотря на очевидную связь между резонансными частотами и частотами пиков амплитудно-частотного спектра, устойчивость и точность определения резонансных частот в большинстве таких методов оказались неудовлетворительными из-за влияния разных факторов (нулей спектра, слияния пиков, присутствия ложных пиков, зависимости от частоты основного тона, различных помех и реверберации помещения). Например, в работах [4–6] исследовался метод мгновенной частоты, который использует оценку интервалов времени между нулями сигналов на выходе гребенки фильтров или их первых и вторых производных по времени. Однако, некоторые варианты этого метода иногда приводят к неправдоподобным результатам, например, к отрицательным значениям формант [7]. Другой метод рассмотрен в [8, 9]. Он использует полюса передаточной функции речевого тракта и позволяет оценить форманты по коэффициентам линейного предсказания. При этом качество получаемых оценок существенно зависит от уровня шумов, особенно в низкочастотной области. Несколько более устойчивым оказался автокорреляционный анализ в частотно-временной области [5].

Можно дать и общую характеристику качества рассмотренных выше подходов к определению формант: ни один из известных методов или их сочетание не гарантируют устойчивости к помехам и погрешность ниже 10% даже для синтезированных сигналов с известными параметрами. Наибольшие трудности при анализе формант реального речевого сигнала связаны с различными типами неустойчивости, которые усугубляются объективно существующими скачками в треках формантных частот.

Альтернативный подход к поиску резонансных частот, связанный с анализом свойств фазовой компоненты комплексного спектра речевого сигнала, был предложен в [10–12]. Это направление исследований развивается в предлагаемой статье и позволяет создать новый метод оценки треков формантных частот, обладающий определенными преимуществами перед упомянутыми выше методами.

2. ТЕОРЕТИЧЕСКИЕ ОСНОВАНИЯ МЕТОДА

Спектрально-временной анализ речевого сигнала $x(t)$ часто выполняется при помощи кратковременного преобразования Фурье

$$S(\omega, t) = \int_0^{\infty} W(t - \tau) e^{-j\omega\tau} x(\tau) d\tau$$

с окном конечной длительности $W(z)$: $W(z) > 0$ для $z \in (0, t_w)$ и $W(z) = 0$ для $z \notin [0, t_w]$. Установлено, что вид и параметры окна сильно влияют на свойства динамического спектра $S(\omega, t)$, особенно при фазовом анализе. При “правильном” выборе формы окна и его длительности t_w в спектре $S(\omega, t)$ проявляются формантные частоты. Однако, при анализе речевых сигналов общего вида нет никаких *физических* оснований для выбора окна, так как спектральные свойства гласных, назальных и фрикативных определяются различными физическими процессами и весьма разнообразны.

Можно попытаться согласовать выбор вида окна со *свойствами восприятия звука* человеком. Ряд психофизиологических экспериментов обнаружил, что в слуховой системе выполняется фильтрация с различными постоянными времени от 2 до 300 мс [12–16]. Кроме того, физиологически обоснованное окно должно учитывать потери в механических структурах слухового анализатора и потери на вязкое трение в гидродинамических процессах в улитке. Однако, этих соображений мало для выбора математической формы окна. В итоге, для каждой конкретной задачи формантного анализа приходится подбирать вид окна и его параметры.

Заметим, что эффект потерь в структурах среднего и внутреннего уха может быть представлен в виде экспоненциально убывающей функции времени. Поэтому рассмотрим такие формы окна, которые учитывают этот эффект и позволят надежно выделять форманты при спектральном анализе. Для этого откажемся от “кратковременности” окна, задавая его зависящим от параметра r в форме

$$W_r(z) = \{e^{-(1+r)z}, z \geq 0; 0, z < 0\}.$$

Тогда получим новое выражение для $S(\omega, t)$:

$$\begin{aligned} S_r(\omega, t) &= \int_0^t e^{-(1+r)(t-\tau)} e^{-j\omega\tau} x(\tau) d\tau = \\ &= \int_0^{\infty} W_r(t - \tau) e^{-j\omega\tau} x(\tau) d\tau. \end{aligned} \quad (1)$$

Непосредственный анализ спектра функции (1) не демонстрирует явных преимуществ перед спектральным анализом функции $S(\omega, t)$ с окна-

ми конечной длительности, однако $S_r(\omega, t)$ играет важную роль в дальнейшем. С помощью величины (1) можно построить новую функцию, для которой формантный анализ оказывается более эффективным, чем при использовании стандартного кратковременного преобразования Фурье.

Вычисление функции $S_r(\omega, t)$ эквивалентно решению при каждом фиксированном ω задачи Коши вида

$$\frac{dS_r(\omega, t)}{dt} = -(1+r)S_r(\omega, t) + x(t)e^{-j\omega t}, \quad S_r(\omega, 0) = 0.$$

Близкое по смыслу уравнение в дискретной форме $y[n] = -ry[n-1] + x[n]e^{-j\omega n}$ с решением $y[n] = y(r, \omega)[n]$ рассматривалось в [17] для $r, 0 < r < 1$. В этой работе была введена новая, дискретная по времени, спектральная функция

$$\alpha_0(r, \omega)[n] = \frac{y(r, \omega + \delta\omega)[n]}{y(r, \omega)[n]}, \quad n = 0, 1, \dots,$$

использующая сдвиг по частоте ω на малую величину $\delta\omega$. Численные эксперименты, выполненные в [18], показали, что когда r близко к 1, фазовый спектр величины $\alpha_0(r, \omega)[n]$ при фиксированном дискретном времени n “похож” на спектр дискретного сигнала $x[n]$. Число r определяет широкополосный или узкополосный тип спектрального анализа. Дискретная величина $y[n] = y(r, \omega)[n]$ является некоторым аналогом непрерывной функции $S_r(\omega, t)$. Заметим, что в работе [17] и в последовавшей затем работе [18] не проводится математическое исследование функции $\alpha_0(r, \omega)[n]$ и обоснование предложенного там алгоритма нахождения формант.

Ниже будет доказано, что использование формально схожего с функцией $\alpha_0(r, \omega)[n]$ выражения

$$\alpha(r, \omega)(t) = \frac{S_r(\omega + \delta\omega, t)}{S_r(\omega, t)}$$

с непрерывным временем t приводит при $r \leq -1$ к созданию эффективного метода анализа формантных частот. Основанием для этого служит подробное математическое исследование свойств величины $\alpha(r, \omega)(t)$. В данном разделе мы приведем лишь краткое изложение полученных результатов, которые подробно описываются в Приложении.

1) При условиях малости $\delta\omega$, а именно при $|\delta\omega|t \ll 1$, $|\delta\omega| \ll \omega$, справедливо следующее представление:

$$\begin{aligned} \alpha(r, \omega)(t) &\approx 1 - j\delta\omega t + j\delta\omega \frac{\int_0^\infty W_{1r}(t-\tau)x(\tau)e^{-j\omega\tau}d\tau}{\int_0^\infty W_r(t-\tau)x(\tau)e^{-j\omega\tau}d\tau} = \\ &= 1 - j\delta\omega \frac{\int_0^t \tau e^{(1+r)\tau} x(\tau)e^{-j\omega\tau}d\tau}{\int_0^t e^{(1+r)\tau} x(\tau)e^{-j\omega\tau}d\tau}, \end{aligned}$$

где $W_{1r}(z) = zW_r(z)$. Условия на $\delta\omega$ считаются выполненными в дальнейшем.

2) Для амплитуды $A_\alpha(r, \omega, t)$ и фазы $\Phi_\alpha(r, \omega, t)$ логарифма величины $\alpha(r, \omega)(t)$ справедливы приближенные (при малых $\delta\omega$) равенства

$$\begin{aligned} A_\alpha(r, \omega, t) &= \ln |\alpha(r, \omega)(t)| \approx -\delta\omega \operatorname{Re}\{U(\omega, r, t)\}, \\ \Phi_\alpha(r, \omega, t) &\approx -\delta\omega \operatorname{Im}\{U(\omega, r, t)\}, \end{aligned}$$

где

$$\begin{aligned} U(\omega, r, t) &= j \left\{ \int_0^t \tau e^{-(1+r)(t-\tau)} x(\tau)e^{-j\omega\tau}d\tau \right\} \times \\ &\times \left\{ \int_0^t e^{-(1+r)(t-\tau)} x(\tau)e^{-j\omega\tau}d\tau \right\}^{-1}. \end{aligned} \quad (2)$$

3) Эти же величины, взятые в конце речевого сегмента, используемого для анализа (формально, при $t \rightarrow \infty$), представляются как

$$\begin{aligned} A(\omega, r) &= \ln |\alpha(r, \omega)(\infty)| \approx -\delta\omega \operatorname{Re}\{U(\omega, r)\}, \\ \Phi(\omega, r) &= \Phi_\alpha(r, \omega, \infty) \approx -\delta\omega \operatorname{Im}\{U(\omega, r)\}, \end{aligned}$$

где

$$\begin{aligned} U(\omega, r) &= j \left\{ \int_0^\infty \tau e^{(1+r)\tau} x(\tau)e^{-j\omega\tau}d\tau \right\} \times \\ &\times \left\{ \int_0^\infty e^{(1+r)\tau} x(\tau)e^{-j\omega\tau}d\tau \right\}^{-1}. \end{aligned}$$

4) Предельные значения величин $A(\omega, r)$, $\Phi(\omega, r)$ при $r \rightarrow -1$, обозначаемые как $A(\omega)$ и $\Phi(\omega)$, выражаются в виде

$$\begin{aligned} A(\omega) &\approx \delta\omega \operatorname{Re} \left\{ -j \frac{\int_0^\infty \tau x(\tau)e^{-j\omega\tau}d\tau}{\int_0^\infty x(\tau)e^{-j\omega\tau}d\tau} \right\} = \\ &= \delta\omega \operatorname{Re} \left\{ \frac{F'[x](\omega)}{F[x](\omega)} \right\} = \delta\omega \operatorname{Re} \{ [\ln F[x](\omega)]' \}, \end{aligned}$$

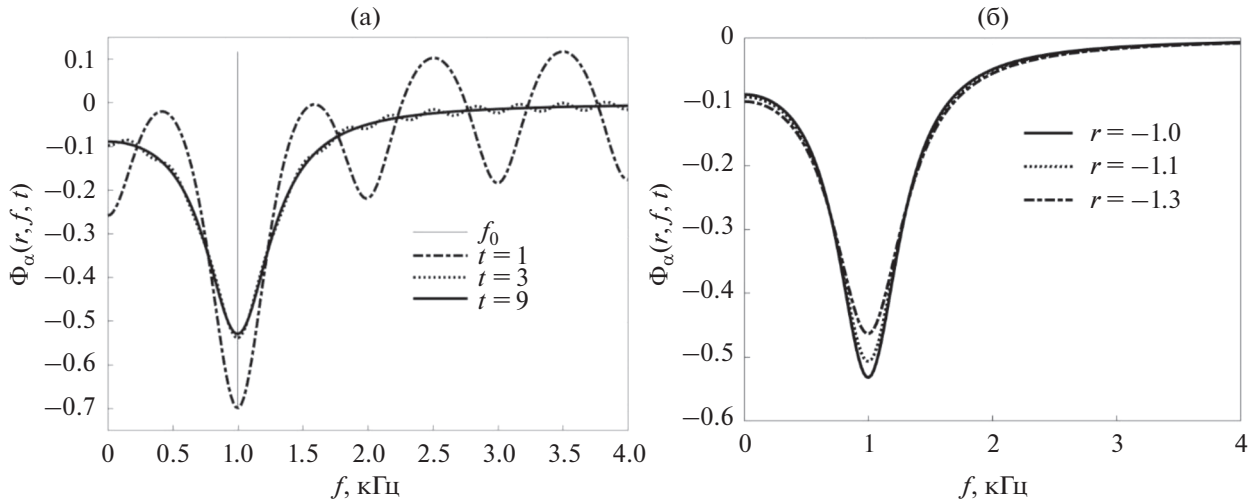


Рис. 1. (а) – Функция $\Phi_\alpha(r, \omega, t)$ при $r = -1.01$ и при различных t . (б) – Функция $\Phi_\alpha(r, \omega, t)$ при “большом” t ($t = 10$) и различных r .

$$\Phi(\omega) \approx \delta\omega \operatorname{Im} \left\{ -j \frac{\int_0^\infty \tau x(\tau) e^{-j\omega\tau} d\tau}{\int_0^\infty x(\tau) e^{-j\omega\tau} d\tau} \right\} =$$

$$= \delta\omega \operatorname{Im} \left\{ \frac{F'[x](\omega)}{F[x](\omega)} \right\} = \delta\omega \operatorname{Im} \{ [\ln F[x](\omega)]' \},$$

где $F[x](\omega) = \int_0^\infty x(\tau) e^{-j\omega\tau} d\tau$ – преобразование Фурье сигнала.

Представленные здесь функции $A_\alpha(r, \omega, t)$, $\Phi_\alpha(r, \omega, t)$, $A(\omega, r)$, $\Phi(\omega, r)$ и $A(\omega)$, $\Phi(\omega)$ приближенно вычисляются по сигналу $x(\tau)$. Компьютерные эксперименты с синтезированными и реальными речевыми сигналами показали, что функции $A(\omega, r)$ и $\Phi(\omega, r)$ можно успешно использовать для нового способа оценки треков резонансных частот речевого тракта. Результаты этих экспериментов представлены ниже в разделах 3 и 4.

3. ЭКСПЕРИМЕНТЫ ПО ОПРЕДЕЛЕНИЮ ФОРМАНТНЫХ ЧАСТОТ

3.1. Формантный анализ модельных сигналов

Рассмотрим поведение фазовой функции $\Phi_\alpha(r, \omega, t)$ для модельного сигнала вида $x(t) = e^{-\sigma t} \sin 2\pi f_0 t$ с частотой f_0 и коэффициентом затухания $\sigma > 0$ при $t \in [0, T_0]$. В этом случае интегралы в формуле (2) вычисляются аналитически, однако из-за громоздкости выражений прямое аналитическое исследование величины $\Phi_\alpha(r, \omega, t)$ как функции частоты $f = \omega/2\pi$ при разных t и r затруднительно.

Можно, однако, провести это исследование численно.

Приведем типичные графики функции $\Phi_\alpha(r, \omega, t)$ при различных t, r для модельных параметров сигнала $f_0 = 1$, $\sigma = 0.08$, $T_0 = 10$. Графики показывают существование хорошо обусловленного минимума фазовой функции при частоте, близкой к “формантной”: $f_0 = 1$. Характерна локализация и стабилизация этого экстремума при увеличении t , т.е. при переходе $\Phi_\alpha(r, \omega, t)$ в функцию $\Phi(\omega, r)$ (рис. 1а) и обострение минимума при фиксированном “большом” t и при $r \rightarrow -1$, т.е. при переходе $\Phi_\alpha(r, \omega, t)$ в функцию $\Phi(\omega)$ (рис. 1б). Заметим, что функцию $\Phi(\omega)$ уже можно исследовать аналитически. Для рассматриваемого модельного сигнала при $\omega = 2\pi f$, $\omega_0 = 2\pi f_0$ получим:

$$F[x](\omega) = \frac{\omega_0}{(\sigma + j\omega)^2 + \omega_0^2},$$

$$(\ln F[x](\omega))' = 2 \frac{\omega - j\sigma}{(\sigma + j\omega)^2 + \omega_0^2},$$

$$\Phi(\omega) \approx 2\delta\omega \operatorname{Im} \left\{ \frac{\omega - j\sigma}{(\sigma + j\omega)^2 + \omega_0^2} \right\} =$$

$$= -2\delta\omega \frac{\omega^2 + \omega_0^2 + \sigma^2}{(\sigma^2 + \omega_0^2 - \omega^2)^2 + 4\sigma^2\omega^2}.$$

Отсюда находится производная функции $\Phi(\omega)$:

$$\Phi'(\omega) = \frac{8\sigma\omega(\sigma^2 + \omega_0^2)[\omega^2 - (\omega_0^2 - \sigma^2)]}{((\sigma^2 + \omega_0^2 - \omega^2)^2 + 4\sigma^2\omega^2)^2}.$$

Из анализа ее знаков ясно, что предельная функция фазы $\Phi(\omega)$ имеет глобальный минимум в точ-

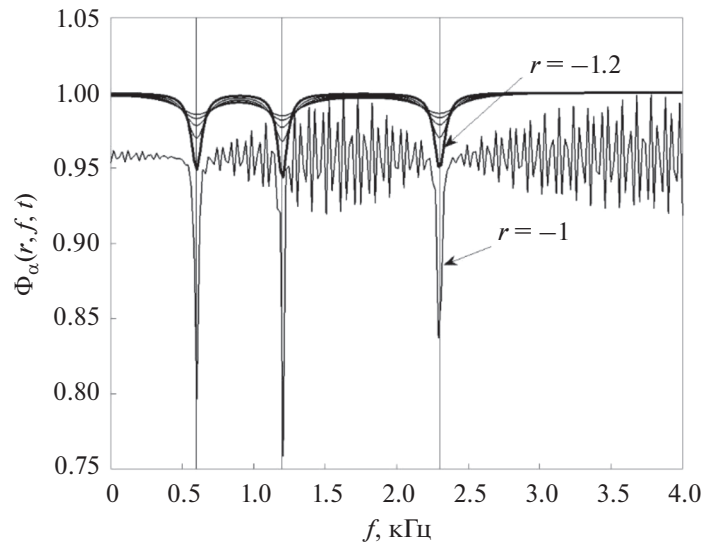


Рис. 2. Функция $\Phi_\alpha(r, \omega, t)$ при для различных r , $r < -1$ и при $t = 10$. Вертикальные линии указывают точные формантные частоты.

ке $\omega^* = \sqrt{\omega_0^2 - \sigma^2}$, так что при “малых” затуханиях ($\sigma/\omega_0 \ll 1$) точка минимума ω^* будет приблизительно совпадать с формантной частотой: $\omega^* \approx \omega_0 - \frac{\sigma^2}{2\omega_0^2} \approx \omega_0$. Эти соображения объясняют поведение функции на рис. 1б.

Аналогичные результаты получаются при численном исследовании модельных сигналов типа суперпозиций нескольких гармоник. Например, для сигнала вида

$$x(t) = A_1 e^{-\delta_1 t} \sin 2\pi f_1 t + A_2 e^{-\delta_2 t} \sin 2\pi f_2 t + A_3 e^{-\delta_3 t} \sin 2\pi f_3 t,$$

$$\begin{aligned} A_1 &= 8, \quad A_2 = -4, \quad A_3 = 1; \\ \delta_1 &= 0.08, \quad \delta_2 = 0.05, \quad \delta_3 = 0.08; \\ f_1 &= 0.6, \quad f_2 = 1.2, \quad f_3 = 2.3, \end{aligned}$$

результат для различных r , $r \leq -1$, при “большом” $t = 10$ показан на рис. 2.

Видно, что формантным частотам отвечают хорошо обусловленные минимумы фазы $\Phi(\omega, r) = \Phi_\alpha(r, \omega, \infty) \approx \Phi_\alpha(r, \omega, 10)$. С увеличением $|r|$ обусловленность минимумов ухудшается, но пропадают побочные минимумы. По этой причине следует находить все минимумы функции $\Phi(\omega, r)$ для различных r , $r \leq -1$ в некотором диапазоне значений вблизи $r = -1$, и анализировать частоту появления различных точек минимума. Проведенные численные эксперименты позволяют предположить, что среди них наиболее часто будут встречаться частоты, близкие к формантным.

Подобные исследования были проведены и для амплитудной функции $A_\alpha(r, \omega, t)$, ее предельного значения $A(r, \omega)$ при $t \rightarrow \infty$ и функции $A(\omega) = A(r = -1, \omega)$. Оказывается, что величины $A(r, \omega), A(\omega)$ обладают свойствами, похожими на свойства функций $\Phi(r, \omega), \Phi(\omega)$. Например, для использованного выше сигнала $x(t) = e^{-\sigma t} \sin 2\pi f_0 t$ функция $A(\omega)$ имеет минимум и максимум соответственно в точках $\omega_\pm(\sigma) = \omega_0(1 \pm \sigma/\omega_0)$, и при малых затуханиях σ эти значения близки к ω_0 . Функция $A(r, \omega)$ при $r < -1$, $r \approx -1$ имеет минимум и максимум в точках $\omega_0(1 \pm \varepsilon)$ при $\varepsilon = \frac{\sigma + |r| - 1}{\omega_0}$, так что формантная частота сигнала близка к значениям точек экстремумов функции $A(r, \omega)$ при $\sigma \rightarrow 0$, $r \rightarrow -1$. Для суперпозиций сигналов типа $x(t)$ с различными частотами f_0 экстремумы функции $A(r, \omega)$ при различных $r < -1$, $r \approx -1$ будут чаще всего встречаться в окрестностях этих частот.

3.2. Синтезированные гласные

Рассмотрим более подробно результаты формантного анализа для синтезированных гласных с известными резонансными частотами. Синтетические речевые сегменты, как и сегменты реальной речи, которые будут использованы в дальнейшем, имеют частоту отсчетов 16 кГц. Характеристики синтезированных гласных соответствуют мужскому голосу с частотой основного тона 120 Гц.

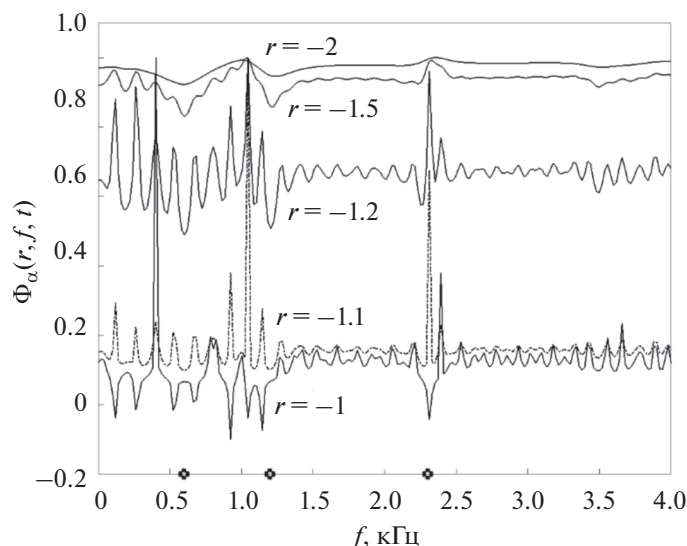


Рис. 3. Функция $\Phi_\alpha(r, \omega, t)$ для синтезированного гласного /а/ для $t = 10$ мс при различных r , $r < -1$. Здесь жирные точки отмечают истинные значения формантных частот.

Проанализируем сегмент синтетической речи, соответствующий звуку /а/ с резонансными частотами $F = \{600, 1200, 2300\}$ Гц, на скользящем сегменте длительностью $T = 10$ мс, так что приближенно можно считать, что $\Phi_\alpha(r, \omega, 10) \approx \Phi(r, \omega)$. На рис. 3 показаны графики вычисленной по такому речевому сигналу функции $\Phi_\alpha(r, \omega, t)$ для различных r , $r \leq -1$, при значении $t = 10$ мс.

Можно видеть, что многочисленные минимумы этой функции для различных r группируются около модельных формантных частот. Поэтому результаты формантного анализа с помощью функции $\Phi(r, \omega)$ целесообразно представлять в виде гистограмм распределения количества минимумов этой функции по частотам. На рис. 4/а/ приведено такое распределение, полученное из графиков рис. 3 для синтетического /а/. На рис. 4/и/, 4/о/, 4/у/ показаны аналогичные распределения для других синтезированных гласных. Рассмотрение этих распределений приводит к выводу о возможности оценки формантных частот по пикам гистограмм минимумов функции $\Phi(r, \omega)$.

Аналогичное исследование было проведено для функции $A_\alpha(r, \omega, t)$. Для получаемой из нее функции $A(r, \omega) = A_\alpha(r, \omega, \infty)$, как отмечалось в предыдущем разделе, информация о резонансных частотах содержится в положении точек ее экстремумов для разных r . Гистограммы распределения этих точек могут также быть источником форматного анализа.

Используя совместно пики гистограмм, получаемых с помощью функций $\Phi(r, \omega)$ и $A(r, \omega)$, можно провести синхронную оценку резонанс-

ных частот на концах периода основного тона. Однако, синхронная оценка требует сегментации речевого сигнала с указанием моментов начала и конца действия голосового источника. Это требует дополнительных вычислительных затрат и вносит определенную погрешность в процедуру форматного анализа. Кроме того, сегментация невозможна для фрагментов речи, где не действует (или слабо действует) голосовой источник (глухие и звонкие фрикативные звуки, сегменты взрыва после смычки, шепотная речь). Вместе с тем, формантный анализ необходим и для таких речевых фрагментов. Поэтому в нашей работе применялся формантный анализ по пикам гистограмм, полученным по функциям $\Phi(r, \omega)$ и $A(r, \omega)$ для сегментов речи с фиксированной длительностью 10 мс, т.е. считалось, что $A(r, \omega) \approx A_\alpha(r, \omega, 10)$ и $\Phi(r, \omega) \approx \Phi_\alpha(r, \omega, 10)$. Пример такого сравнительного анализа гистограмм и оценки формантных частот для синтетического звука /а/ приведен на рис. 5. Из рис. 5 видно, что пики распределения минимумов функции $\Phi(r, \omega)$ достаточно близки к формантным частотам звука. При этом распределение экстремумов функции $A(r, \omega)$ позволяет надежно оценить далеко не все форманты.

Средние значения формантных частот для каждого из рассмотренных синтетических гласных, найденные по распределениям минимумов функции $\Phi(r, \omega)$, представлены в табл. 1. Известные резонансные частоты гласных выделены жирным шрифтом. Погрешность оценок резонансных частот оказалась достаточно малой, особенно для резонансов выше первого, на значения которого влияет частота основного тона.

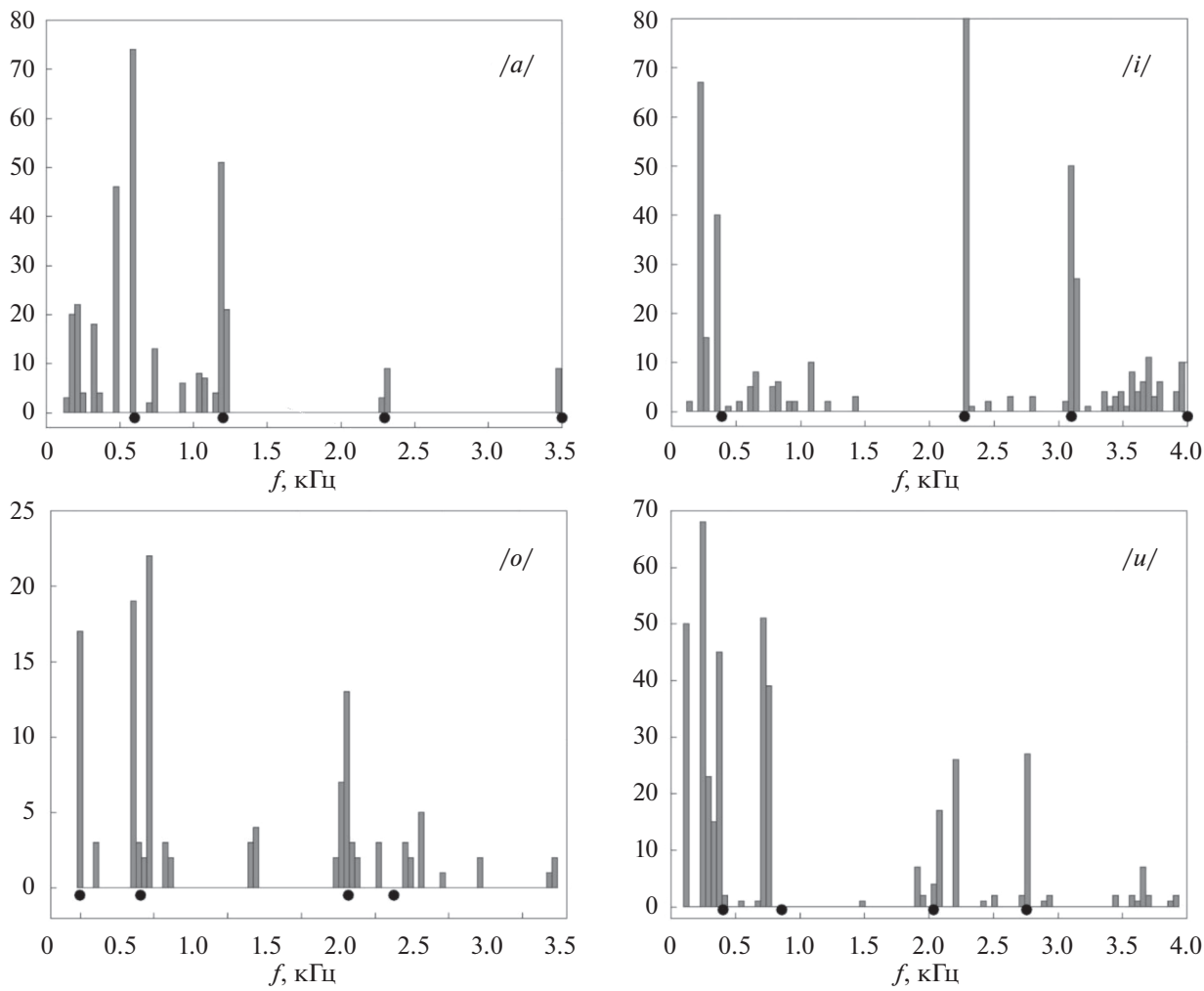


Рис. 4. Гистограммы распределений точек минимума функции $\Phi_\alpha(r, \omega, t)$ при $t = 10$ для различных синтезированных гласных. Черными точками обозначены истинные значения формантных частот.

Вычисляя пики гистограмм распределения минимумов функции $\Phi(r, \omega)$ для каждого периода основного тона сегмента речи (или для каждого ее последовательного временного отрезка фиксиро-

ванной длины), можно проследить динамику изменения формантных частот от периода к периоду. В результате получаем *фазограмму*, сформированную с помощью фазового анализа речевого

Таблица 1. Исходные и вычисленные по минимумам $\Phi(r, \omega)$ резонансные частоты синтезированных гласных при отсутствии шума в сигнале

	А	Э	О	У	И	Ы
F_1 , Гц	600 585	565 562	497 545	309 345	290 328	286 328
F_2 , Гц	1200 1265	1381 1430	914 955	758 782	2272 2269	1874 1899
F_3 , Гц	2300 2284	2252 2298	2316 2347	2042 2094	3100 3194	2575 2598
F_4 , Гц	3500 3558	2789 3326	2625 2593	2761 2793	4000 4000	3732 3735

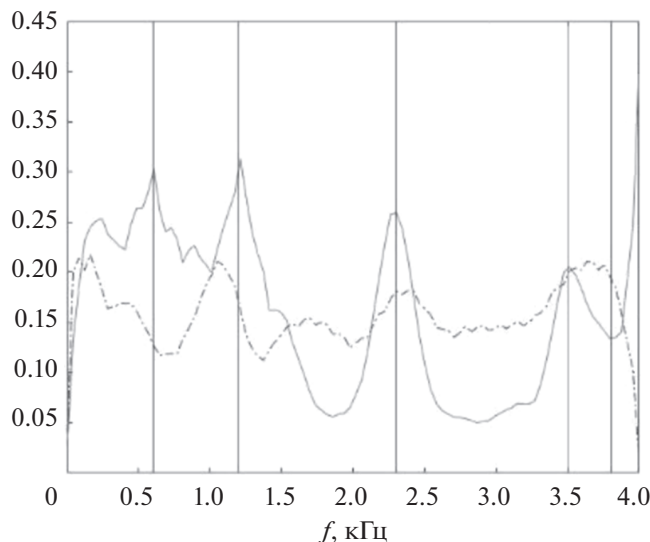


Рис. 5. Оценка формантных частот по распределению минимумов функции $\Phi(r, \omega)$ (—) и распределению экстремумов функции $A(r, \omega)$ (---). Вертикальными линиями отмечены резонансные частоты гласного /a/.

сигнала по функции $\Phi(r, \omega)$. Соответственно, с помощью распределений экстремумов функции $A(r, \omega)$ на каждом периоде получаем аналог спектрограммы сигнала. Примеры фазограмм для синтетического звука /a/ приведены на рис. 6.

Было также проведено исследование полученных оценок резонансных частот на устойчивость

к возмущениям сигналов. Для этого к сигналу добавлялся случайный белый шум с разными отношениями сигнал/шум (SNR). Качественно степень устойчивости можно оценить, сравнивая визуально фазограммы синтетического гласного /a/ без добавления шума (рис. 6а) и с шумом при SNR = 10 дБ (рис. 6б). Кроме того, на рис. 6в, 6г показаны соответствующие динамические оценки пяти резонансных частот, полученные по пикам распределений минимумов функции $\Phi(r, \omega)$.

Благодаря стационарности формант синтезированных гласных, можно количественно оценить устойчивость оценок частот, найденных по функциям $\Phi(r, \omega)$ и $A(r, \omega)$, и их зависимость от уровня шума. Соответствующие сравнительные данные об ошибках вычисления резонансных частот представлены в табл. 2, где погрешность для сигнала без шума представлена жирным шрифтом. Из таблицы можно, например, видеть, что для гласного /a/ оценки по $\Phi(r, \omega)$ стабильны вплоть до третьей форманты, тогда как оценки частот по функции $A(r, \omega)$ демонстрируют заметный разброс. За редкими исключениями, погрешность оценок не сильно возрастает и с увеличением уровня шума вплоть до SNR = 5 дБ.

3.3. Реальные речевые сигналы

Теперь приведем результаты использования предлагаемой нами процедуры формантного анализа для реальных речевых сигналов. В целом они

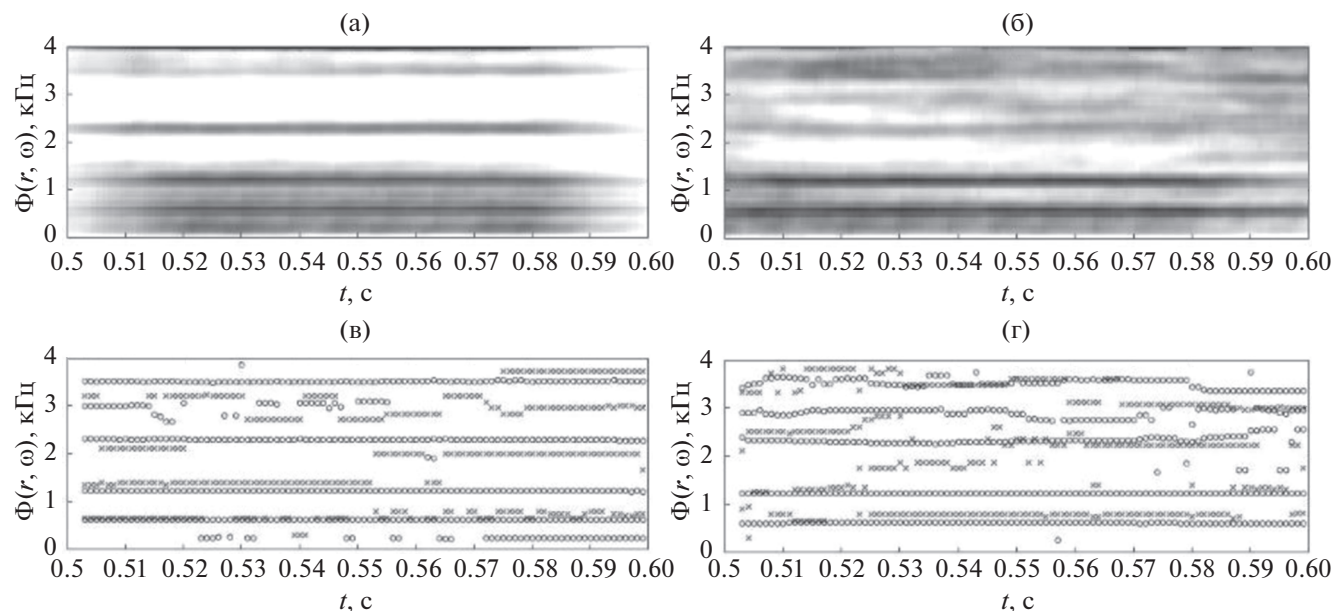


Рис. 6. Формантный анализ синтетического гласного /a/. (а), (в) — Сигнал без шума и (б), (г) — смесь его со случайным шумом при SNR = 10 дБ. (а), (б) — Фазограммы, найденные по функции $\Phi(r, \omega)$; (в), (г) — динамические оценки (треки) резонансных частот гласного по пикам гистограмм минимумов функции $\Phi(r, \omega)$ (ooo) и экстремумов функции $A(r, \omega)$ (xxx).

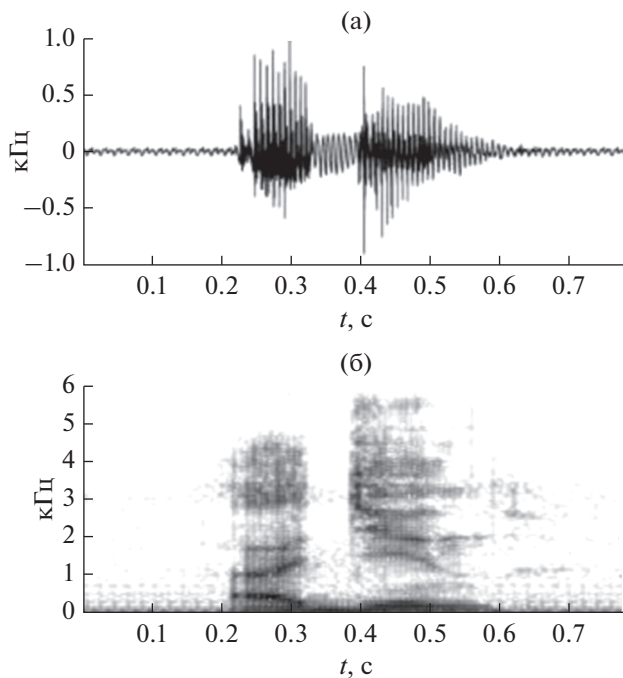
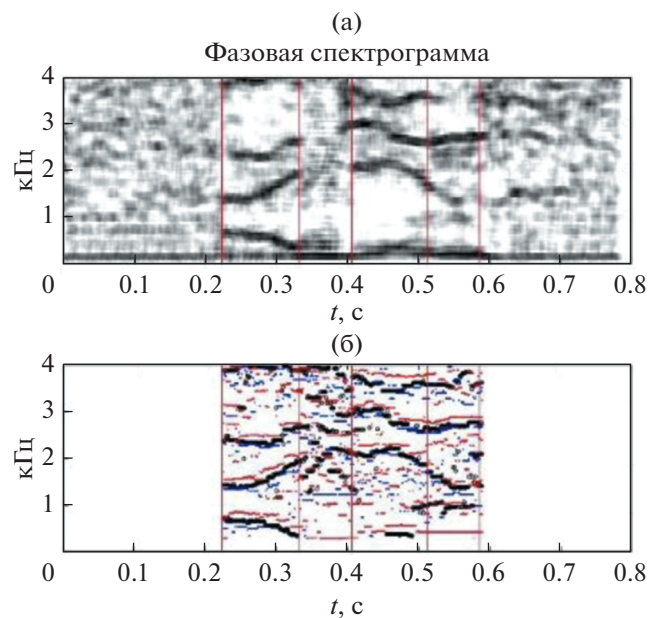
Таблица 2. Ошибки оценки резонансных частот εF_i , % (SNR = 5 дБ)

	А	Э	О	У	И	Ы
εF_1 , Гц	2.5	0.5	-9.6	-11.6	-13.1	-14.6
SNR = 5 дБ	2.3	3.5	9.6	12.3	13.1	14.7
εF_2 , Гц	-5.4	-3.5	-4.4	-3.1	0.1	-1.3
SNR = 5 дБ	-6.1	2.3	-9.1	-6.4	0.1	-1.3
εF_3 , Гц	0.7	-2.0	-1.3	-2.5	-3.0	-0.9
SNR = 5 дБ	-0.2	-1.4	3.5	-7.1	-6.0	-6.1
εF_4 , Гц	-1.6	-19.2	1.2	-1.1	0.0	-0.1
SNR = 5 дБ	14.5	2.9	1.4	-2.5	10.9	-0.4

подтверждают, по меньшей мере качественно, возможность выделения треков формантных частот. Однако, обнаруживается также нестабильность оценок и их отсутствие в некоторые моменты времени. Рассмотрим, например, сигнал для слова /один/ (см. рис. 7а). По сравнению с его стандартной сонограммой (рис. 7б), треки формант, полученные с помощью функции $\Phi(r, \omega)$ (рис. 8а), выделяются более четко, в том числе и вторая форманта, возникающая в момент опускания небной занавески перед назальной смычкой вблизи момента времени 0.5 с. При этом мгновенная оценка пиков в каждый момент времени без последующей фильтрации иногда приводит к множеству ложных значений (рис. 8б). Поэтому фильтрация ложных оценок необходима, как и

сглаживание треков достоверных оценок резонансов. Это, впрочем, характерно для любого метода поиска резонансных частот речевого сигнала. В этом и последующих экспериментах экстремумы $\Phi(r, \omega)$ сортировались по убыванию величины, и первые 5 оценок частоты принимались за формантные частоты.

Оценка формантных частот по экстремумам функции $A(r, \omega)$ в некоторые моменты времени совпадает с оценками по минимумам $\Phi(r, \omega)$, но имеются моменты, где они оказываются завышенными. Необходимо также отметить, что на интервале звонкой смычки /д/ в слове /один/ (на сегменте 0.35–0.4 с) экстремумы $\Phi(r, \omega)$ приходятся на предположительно радиальный резо-

**Рис. 7.** Слово /один/. (а) — Осциллограмма, (б) — сонограмма. Диапазон частот 0–6 кГц.**Рис. 8.** Слово /один/. (а) — Фазограмма, (б) — треки пиков. Оценки по $\Phi(r, \omega)$ (○); по $A(r, \omega)$ (●). Диапазон частот 0–4 кГц.

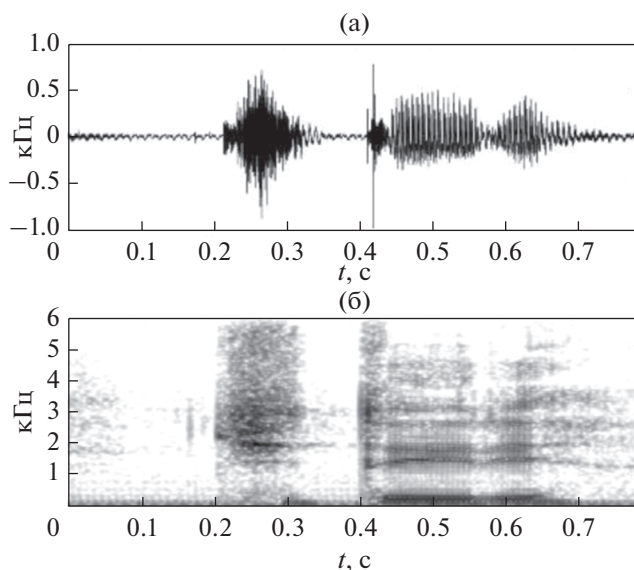


Рис. 9. Слово /четыре/. (а) — Осциллограмма, (б) — сонограмма. Диапазон частот 0–6 кГц.

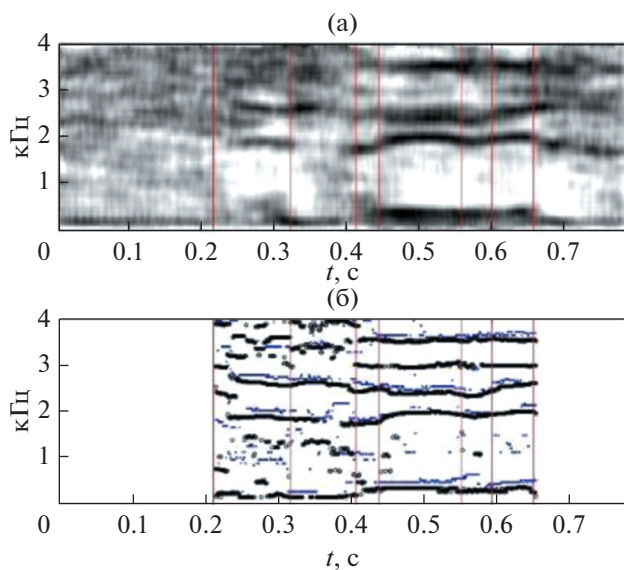


Рис. 10. Слово /четыре/. (а) — Фазограмма, (б) — треки пиков. Оценки по $\Phi(r, \omega)$ (ooo); по $A(r, \omega)$ (●●●). Диапазон частот 0–4 кГц.

нанс в области 0.2 кГц, который косвенно указывает на место артикуляции.

Вертикальные линии на рис. 8 (и далее, на рис. 10 и 12) обозначают границы фонетических сегментов, которые были заданы вручную.

На рис. 9 и 10 показаны осциллограмма, сонограмма и фазограмма, а также треки пиков распределения минимумов функции $\Phi(r, \omega)$ для слова /четыре/, произнесенного с выпадением звука /е/ после фрикативного /ч/ и редукцией конечного

звука /е/, так что фонетическая транскрипция этого слова есть /чтыри/. На сегменте взрыва глухой смычки 0.413–0.445 с присутствует оценка второй, третьей и четвертой форманты, что необходимо для распознавания места артикуляции согласного /т/. Также хорошо выражена девиация треков на сегменте 0.558–0.6 с, соответствующая артикуляции звука /р/. Кроме того, представлены треки на глухом фрикативном /ч/, продолжающиеся на глухой смычке вследствие реверберации.

Как показано на рис. 8 и 10, формантные частоты могут определяться и на сегментах с шумовым или импульсным возбуждением, и это представляется важным преимуществом описываемого подхода. Можно также сравнить сонограммы и фазограммы слова /один/, произнесенного обычным голосом (рис. 7 и 8) и шепотом (рис. 11 и 12) тем же диктором. Видно, что формантные треки шепотных гласных достаточно хорошо выражены, хотя и наблюдаются некоторые отличия от тех, что показаны на рис. 7 и 8. Это отличие вызвано, главным образом, различием в акустике речеобразования с голосовым и шепотным возбуждением.

Итак, эксперименты с синтезированными и реальными речевыми сигналами подтверждают, что гистограммы минимумов функции $\Phi(r, \omega)$ и экстремумов функции $A(r, \omega)$ содержат информацию о резонансных частотах речевого сигнала. Положения пиков этих гистограмм можно использовать для формантного анализа.

4. ОБСУЖДЕНИЕ

В данной работе использовался спектральный анализ фазы $\Phi(r, \omega)$ и амплитудной функции $A(r, \omega)$ величины $\alpha(r, \omega)(t)$. Количественная оценка погрешности определения частоты формант этим методом была выполнена в экспериментах с синтезированными гласными звуками. Говоря о качестве такого анализа, нужно иметь в виду, что оценка эффективности любого метода поиска формантных частот только применительно к синтезированным гласным весьма условна в силу упрощенной структуры таких сигналов по сравнению с реальными речевыми сегментами. Однако, наш анализ и численные эксперименты подтверждают возможность определения резонансных частот с помощью функций $\Phi(r, \omega)$ и $A(r, \omega)$. Эксперименты с синтезированными гласными позволяют также оценить устойчивость найденных формант к аддитивным помехам. В наших экспериментах ошибка вычисленных формантных частот по возмущенным данным оказалась вполне приемлемой для практических приложений вплоть до весьма высокого уровня шумов с отношением сигнал/шум +5 дБ. Это позволяет надеяться на то, что и для речевых сигналов в ре-

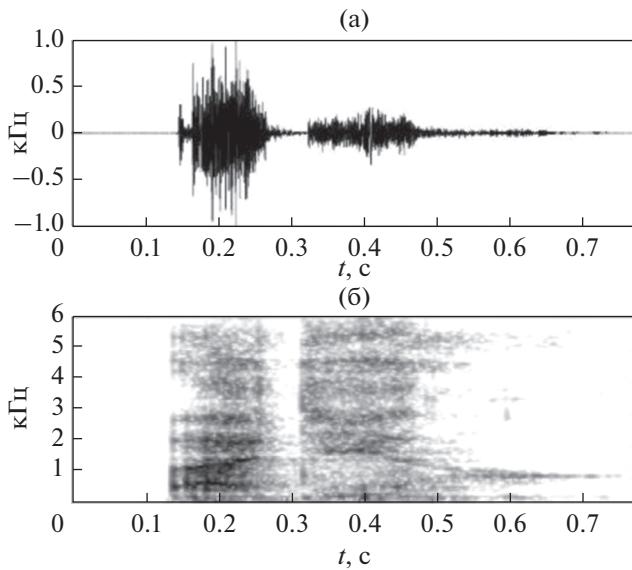


Рис. 11. Слово /odin/, шепот. (а) — Осциллограмма, (б) — сонограмма.

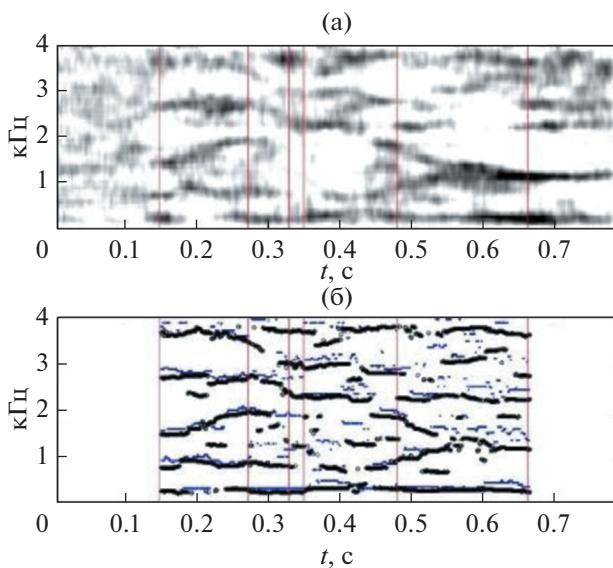


Рис. 12. Слово /odin/, шепот. (а) — Фазограмма, (б) — треки пиков. Оценки по $\Phi(r, \omega)$ (ooo); по $A(r, \omega)$ (●●●). Диапазон частот 0–4 кГц.

альной акустической обстановке формантный анализ предложенным методом окажется достаточно устойчивым к шумам и другим артефактам. Здесь же следует отметить, что результаты нашей процедуры формантного анализа не зависят от амплитуды исследуемого сигнала. В результате достигается инвариантность к текущему уровню его громкости. Вместе с тем, такая инвариантность затрудняет фильтрацию шума на паузах и смычках.

Необходимо отметить важное преимущество используемого в данной статье метода. Оно состоит в возможности оценки формантных частот в любой момент времени, независимо от типа источника возбуждения акустических колебаний в речевом тракте. Это выгодно отличается от синхронного метода и оценки по групповой задержке [11], где нужно предварительно найти моменты открытия голосовой щели на периоде основного тона. Ошибка в определении этого момента может привести к большой погрешности или к невозможности получить оценки формантных частот. В предлагаемом методе такие ситуации исключены. Отметим также, что параллельная оценка формантных частот по гистограммам минимумов функций $\Phi(r, \omega)$ и экстремумов функции $A(r, \omega)$ полезна как источник дополнительной информации, несмотря на то, что последняя оказывается менее надежной.

Предлагаемый метод можно также применить для решения следующей проблемы. В любом методе нахождения формант необходимо выяснить, какие элементы их треков являются оценками формант ротовой, носовой или подсвязочной области, а какие — артефактами. Один из способов состоит в агрегировании оценок, полученных разными методами как в амплитудно-частотной, так и в фазо-частотной области. Подход подобного рода можно распространить и на оценки формантных частот по функциям $\Phi(r, \omega)$ и $A(r, \omega)$ на сегментах взрыва смычных согласных с учетом последующего гласного.

В конце обсуждения выскажем гипотезу об отношении величины $\alpha(r, \omega)(t)$, лежащей в основе нашего анализа, к детекторам частотных модуляций, найденным в слуховой системе (см., например, [19]). Один из вариантов математического описания таких детекторов связан с величиной $A(\omega, t) = \lg[S(\omega + \delta\omega, t)/S(\omega, t) + c]$, схожей по форме с $\alpha(r, \omega)(t)$. Эта аналогия позволяет предположить, что использованная нами математическая модель имеет отношение к свойствам слухового анализатора человека, а найденные формантные частоты близки к частотам, оцениваемым этим анализатором.

ЗАКЛЮЧЕНИЕ

Сформулируем основные результаты этой статьи.

1) Предложен новый метод оценки динамических треков формантных частот для произвольных речевых сегментов. Он использует отношение $\alpha(r, \omega)(t)$ двух преобразований Фурье речевого сигнала со специальными параметрическими окнами экспоненциального типа $W_r(z)$, $W_{lr}(z)$. Это отношение рассматривается как функция частоты и параметра r . Анализируя для нескольких

значений параметра r , $r < -1$ распределение точек минимума (по частоте) для фазы отношения $\alpha(r, \omega)(t)$ и/или аналогичное распределение точек экстремума для его амплитуды, можно по получаемым пикам распределений оценить формантные частоты.

2) Представлено математическое обоснование такого подхода к оценкам формантных частот для простейших модельных сигналов.

3) Проведены большие серии численных экспериментов по обработке синтетических и реальных речевых сигналов, подтвердившие практическую работоспособность предложенного метода оценки формант. Установлено, что оценка по фазе отношения более надежна, чем по амплитуде. Вместе с тем, оценка по амплитуде может быть использована как источник дополнительной информации о формантах.

4) В экспериментах с синтезированными гласными было установлено, что погрешность оценки их резонансных частот мала и устойчива по отношению к аддитивным шумам вплоть до отношения сигнал/шум 5 дБ.

5) Для сегментов реальной речи показано, что метод позволяет вычислить треки формантных частот как для звуков с голосовым возбуждением, так и для глухих фрикативных, аспиративных взрывов и шепотной речи.

При выполнении работы первый автор пользовался поддержкой Программы повышения конкурентоспособности Национального исследовательского ядерного университета МИФИ (Московского инженерно-физического института), проект № 02.а03.21.0005 от 27.08.2013.

ПРИЛОЖЕНИЕ

Здесь представлен вывод формул для функций $A(\omega, r)$, $\Phi(\omega, r)$, а также $A(\omega)$ и $\Phi(\omega)$, применяемых в предлагаемом методе формантного анализа.

Введем обозначение $y(r, \omega)(t) = S_r(\omega, t)$, где функция $S_r(\omega, t)$ определяется формулой (1). Такое обозначение подчеркивает, что эта величина зависит от времени, а от величин r, ω она зависит как от параметров. Кроме частоты ω , будем использовать “близкие” частоты $\omega + \delta\omega$ ($|\delta\omega| \ll \omega$). Определим отношение

$$\alpha(r, \omega)(t) = \frac{y(r, \omega + \delta\omega)(t)}{y(r, \omega)(t)} = 1 + \frac{\Delta y}{y},$$

$$\Delta y = y(r, \omega + \delta\omega) - y(r, \omega).$$

Далее, с учетом “малости” величины $\delta\omega$ и формулы (1) получим:

$$\begin{aligned} \Delta y &= y(r, \omega + \delta\omega) - y(r, \omega) \approx \frac{\partial y}{\partial \omega}(r, \omega)\delta\omega = \\ &= \delta\omega \int_0^\infty W_r(t - \tau)x(\tau)(-j\tau)e^{-j\omega\tau}d\tau = \\ &= j\delta\omega \int_0^\infty W_r(t - \tau)x(\tau)(t - \tau)e^{-j\omega\tau}d\tau - \\ &- j(\delta\omega)t \int_0^\infty W_r(t - \tau)x(\tau)e^{-j\omega\tau}d\tau = \\ &= j\delta\omega \left(\int_0^\infty W_{1r}(t - \tau)x(\tau)e^{-j\omega\tau}d\tau - \right. \\ &\left. - t \int_0^\infty W_r(t - \tau)x(\tau)e^{-j\omega\tau}d\tau \right), \end{aligned}$$

где $W_{1r}(z) = zW_r(z)$. Отсюда

$$\begin{aligned} \alpha(r, \omega)(t) &= 1 + \frac{\Delta y}{y} \approx 1 - j\delta\omega t + \\ &+ j\delta\omega \frac{\int_0^\infty W_{1r}(t - \tau)x(\tau)e^{-j\omega\tau}d\tau}{\int_0^\infty W_r(t - \tau)x(\tau)e^{-j\omega\tau}d\tau} = \\ &= 1 - j\delta\omega t + j\delta\omega \frac{\int_0^t (t - \tau)e^{-(1+r)(t-\tau)}x(\tau)e^{-j\omega\tau}d\tau}{\int_0^t e^{-(1+r)(t-\tau)}x(\tau)e^{-j\omega\tau}d\tau} = \\ &= 1 - j\delta\omega \frac{\int_0^t \tau e^{-(1+r)(t-\tau)}x(\tau)e^{-j\omega\tau}d\tau}{\int_0^t e^{-(1+r)(t-\tau)}x(\tau)e^{-j\omega\tau}d\tau}. \end{aligned}$$

В дальнейшем нас будут интересовать величина $A_\alpha(r, \omega, t) = \ln |\alpha(r, \omega)(t)|$ и фаза $\Phi_\alpha(r, \omega, t)$ отношения $\alpha(r, \omega)(t)$. Мы будем для краткости называть эти величины амплитудной и фазовой функциями, связанными с отношением $\alpha(r, \omega)(t)$. Для их нахождения используем формулу

$$\begin{aligned} \ln \alpha(r, \omega)(t) &= \ln |\alpha(r, \omega)(t)| + j\Phi_\alpha(r, \omega, t) = \\ &= A_\alpha(r, \omega, t) + j\Phi_\alpha(r, \omega, t). \end{aligned}$$

Тогда при условиях $|\delta\omega|t \ll 1$, $|\delta\omega| \ll \omega$, т.е. при $\left| \frac{\Delta y}{y} \right| \ll 1$, имеем:

$$\begin{aligned} \ln \alpha(r, \omega)(t) &= \ln \left(1 + \frac{\Delta y}{y} \right) \approx \frac{\Delta y}{y} \Rightarrow \\ \Rightarrow A_\alpha(r, \omega, t) &\approx \operatorname{Re} \left\{ \frac{\Delta y}{y} \right\}, \quad \Phi_\alpha(r, \omega, t) \approx \operatorname{Im} \left\{ \frac{\Delta y}{y} \right\}. \end{aligned}$$

Отсюда

$$A_{\alpha}(r, \omega, t) \approx -\delta\omega \operatorname{Re} \left\{ j \frac{\int_0^t \tau e^{-(1+r)(t-\tau)} x(\tau) e^{-j\omega\tau} d\tau}{\int_0^t e^{-(1+r)(t-\tau)} x(\tau) e^{-j\omega\tau} d\tau} \right\},$$

$$\Phi_{\alpha}(r, \omega, t) \approx -\delta\omega \operatorname{Im} \left\{ j \frac{\int_0^t \tau e^{(1+r)\tau} x(\tau) e^{-j\omega\tau} d\tau}{\int_0^t e^{(1+r)\tau} x(\tau) e^{-j\omega\tau} d\tau} \right\}.$$

Амплитудная и фазовая функции, связанные с $\alpha(r, \omega)(t)$, для ограниченного по амплитуде сигнала $x(t)$ стабилизируются при $t \rightarrow \infty$, т.е., как мы будем далее говорить, в конце речевого сегмента:

$$A(\omega, r) = \ln |\alpha(r, \omega)(\infty)| = -\delta\omega \operatorname{Re} \{U(\omega, r)\}, \quad (П1)$$

$$\Phi(\omega, r) = \Phi_{\alpha}(r, \omega, \infty) = -\delta\omega \operatorname{Im} \{U(\omega, r)\},$$

где

$$U(\omega, r) =$$

$$= j \left\{ \int_0^{\infty} \tau e^{(1+r)\tau} x(\tau) e^{-j\omega\tau} d\tau \right\} \left\{ \int_0^{\infty} e^{(1+r)\tau} x(\tau) e^{-j\omega\tau} d\tau \right\}^{-1}. \quad (П2)$$

Выражения (П1) и (П2) будут использованы для формантного анализа речевых сигналов общего вида $x(t)$. Предпосылкой для этого служит численное исследование их поведения при различных $r \leq -1$ для сигналов простого вида (см. раздел 3), а также аналитическое исследование их предельных значений $A(\omega)$ и $\Phi(\omega)$ при $r \rightarrow -1$, т.е. величин

$$A(\omega) = A_{\alpha}(-1, \omega, \infty) \approx \delta\omega \operatorname{Re} \{U(\omega, -1)\} =$$

$$= \delta\omega \operatorname{Re} \left\{ -j \frac{\int_0^{\infty} \tau x(\tau) e^{-j\omega\tau} d\tau}{\int_0^{\infty} x(\tau) e^{-j\omega\tau} d\tau} \right\} = \delta\omega \operatorname{Re} \left\{ \frac{F'[x](\omega)}{F[x](\omega)} \right\},$$

$$\Phi(\omega) = \Phi_{\alpha}(-1, \omega, \infty) \approx \delta\omega \operatorname{Im} \{U(\omega, -1)\} =$$

$$= \delta\omega \operatorname{Im} \left\{ \frac{F'[x](\omega)}{F[x](\omega)} \right\} = \delta\omega \operatorname{Im} \{[\ln F[x](\omega)]\}.$$

Здесь $F[x](\omega) = \int_0^{\infty} x(\tau) e^{-j\omega\tau} d\tau$ — преобразование Фурье сигнала. Это исследование представлено в разделе 3.

СПИСОК ЛИТЕРАТУРЫ

1. Stevens K.N. Acoustic Phonetics. The MIT Press, 1998.
2. Sorokin V.N., Leonov A.S., Makarov I.S., Tsyplikhin A.I. Speech inversion and re-synthesis // InterSpeech. 2005. P. 3209–3212.
3. Сорокин В.Н. Речевые процессы. М.: Народное образование, 2012.
4. Sreenivas Th., Niederjohn R.J. Zero-crossing based spectral analysis and SVD spectral analysis for formant frequency estimation in noise // IEEE Trans. Signal Processing. 1992. V. 40. № 2. P. 282–293.
5. Сорокин В.Н., Трифоненков В.П. Об автокорреляционном анализе речевого сигнала // Акуст. журн. 1996. Т. 4. № 3. С. 418–425.
6. Леонов А.С., Макаров И.С., Сорокин В.Н. Частотные модуляции в речевом сигнале // Акуст. журн. 2009. Т. 55. № 6. С. 809–821.
7. Vakman D. On the Analytic Signal, the Teager-Kaiser Energy Algorithm, and Other Methods for Defining Amplitude and Frequency // IEEE Trans. Signal Processing. 1996. V. 44. № 4. P. 791–797.
8. Маркел Дж.Д., Грей А.Х. Линейное предсказание речи. М.: Связь, 1980.
9. Yegnanarayana B. Formant extraction from linear-prediction phase spectra // J. Acoust. Soc. Am. 1978. V. 63. P. 1638–1640.
10. Леонов А.С., Сорокин В.Н. Формантный анализ в фазовой области // Информационные процессы. 2021. Т. 21. № 2. С. 125–134.
11. Сорокин В.Н., Леонов А.С. Фазовые модуляции в речевом сигнале // Акуст. журн. 2022. Т. 68. № 2. С. 218–232.
12. Raab D.H. Forward and backward masking between acoustic clicks // J. Acoust. Soc. Am. 1961. V. 33. С. 137–139.
13. Elliot L.L. Backward and forward masking of probe tones of different frequencies // J. Acoust. Soc. Am. 1962. V. 34. P. 1116–1117.
14. Babkoff H., Sutton S. Monaural temporal masking of transients // J. Acoust. Soc. Am. 1968. V. 44. P. 1373–1378.
15. Green D.M. Temporal acuity as a function of frequency // J. Acoust. Soc. Am. 1973. V. 54. P. 373–379.
16. Hermansky H., Morgan N. RASTA processing of speech // IEEE Trans. Speech and Audio Processing. 1994. V. 2. № 4. P. 578–589.
17. Yegnanarayana B. Group delay spectrogram of speech signals without phase wrapping // J. Acoust. Soc. Am. 2022. V. 151. № 3. P. 2181–2191.
18. Yegnanarayana B. Analysis of phase derivatives of speech signals // J. Acoust. Soc. Am. 2022. V. 152. № 3. P. 1721–1736.
19. Сорокин В.Н., Ченелев Д.Н. Первичный анализ речевых сигналов // Акуст. журн. 2005. Т. 51. № 4. С. 536–542.