АКУСТИЧЕСКИЙ ЖУРНАЛ, 2009, том 55, № 6, с. 809–821

АКУСТИКА ЖИВЫХ СИСТЕМ. БИОЛОГИЧЕСКАЯ АКУСТИКА

УДК 612.85

ЧАСТОТНЫЕ МОДУЛЯЦИИ В РЕЧЕВОМ СИГНАЛЕ

© 2009 г. А. С. Леонов, И. С. Макаров*, В. Н. Сорокин*

Московский инженерно-физический институт

115409 Москва, Каширское ш. 31

** Институт проблем передачи информации РАН*

127994 Москва, Б. Каретный пер. 19

E-mail: vns@iitp.ru

Поступила в редакцию 30.09.08 г.

Исследуются физические механизмы частотных модуляций в акустике речевого тракта и методы оценки этих модуляций в речевом сигнале. Установлено, что колебания стенок тракта оказывают пренебрежимо малое влияние на модуляции его резонансных частот. Модель процесса речеобразования, учитывающая подсвязочную область, показывает, что изменение граничных условий при открытой голосовой щели создает заметные вариации резонансных частот. Наряду с модуляциями такого рода, в речевом сигнале возникают и модуляции, обусловленные влиянием формы источника возбуждения. Они существенно зависят от соотношения частоты основного тона и резонансной частоты, а также от параметров методов оценки модуляций и метода анализа речевого сигнала. В целом, это иногда может привести к нестабильным и непредсказуемым модуляциям вычисленных формантных частот в речевом сигнале.

PACS: 43.72.Pf, 43.72.Fx

1. ВВЕДЕНИЕ

Для решения многих речевых обратных задач нужна точная и надежная оценка резонансных частот (формант) речевого тракта в зависимости от времени. Эти зависимости обычно носят сложный характер, связанный с наложением быстрых колебаний (модуляций) на сравнительно медленные изменения формант. Медленные изменения резонансных частот определяются движениями органов речевого тракта. Быстрые модуляции оказываются синхронными с колебаниями голосовой щели. При решении обратной задачи о нахождении формы речевого тракта по трекам формант для звуков с голосовым возбуждением необходимо вычислять резонансные частоты на таких временных интервалах, где влияние голосового возбуждения наименьшее, и основную роль играет именно форма тракта. В связи с этим важно выяснить механизм появления модуляций формант с тем, чтобы отделить эти модуляции от формантных треков.

Явление частотной модуляции в речевом сигнале – это экспериментально установленное свойство. Его, например, можно наблюдать, регистрируя изменение формы спектра речевого сигнала и степень выраженности в нем формантных частот при сдвиге окна анализа относительно импульса голосового возбуждения.

В литературе описаны различные методы оценки модуляций формантных частот. В работе [1] формантные колебания выделяются фильтрами Габора, после чего мгновенная частота по отклику каждого фильтра определяется с помощью оператора разделения энергии Тигера – Кайзера. В [2] результаты этого алгоритма сопоставляются с оценкой мгновенной частоты с помощью преобразования Гильберта, и делается вывод о близости этих методов. Дальнейшие обобщения алгоритма разделения энергии содержатся в [3]. В работах [4, 5] мгновенная частота оценивается с помощью ковариационного метода линейного предсказания второго порядка. Алгоритм, основанный на использовании адаптивных нуль-полюсных фильтров и линейного предсказания в частотной области, построен в [6, 7]. В работах [8, 9] частотные модуляции определяются с помощью анализа нулей откликов полосовых фильтров. В [10, 11] для оценки мгновенной частоты используются нули некоторых функций, описывающих амплитуду и фазу формантных колебаний. В [12] амплитудные и частотные модуляции определяются с помощью итеративного алгоритма, основанного на преобразовании Гильберта.

Несмотря на разнообразие методов, вопрос о физической адекватности получаемых с их помощью оценок мгновенной частоты остается открытым. В [13] построен пример ошибочного определения частотной модуляции с помощью алгоритма разделения энергии. При этом разница между истинной мгновенной частотой и частотой, оцененной этим алгоритмом, оказалась чрезвычайно большой. В [7] указывалось на возможность полу-

Таблица 1. Вариации первой резонансной частоты по [17]

Гласная	Закрытая голосовая щель, F_1, Гц	Площадь голосовой щели 0.08 см^2 F_1, Гц	Девиация %	Площадь голосовой щели 0.12 см^2 F_1, Гц	Девиация %
A	677	806	+19	858	+26.7
E	459	475	+3.5	482	+5.0
O	538	582	+8.2	582	+8.2
U	291	308	+5.8	323	+11.0
I	285	297	+4.2	305	+7.0

чения отрицательных значений мгновенной частоты с помощью алгоритмов из [1, 2].

Другим мало исследованным вопросом является проблема устойчивости оценок мгновенной частоты относительно внешних шумов и различных типов микрофона. В [1] к тестовым сигналам примешивался белый шум с различным отношением "сигнал/помеха". Выяснилось, что при отношении 30 дБ среднеквадратическая погрешность оценки мгновенной частоты составила около 10%. При уменьшении отношения "сигнал/помеха" до 20 дБ среднеквадратическая погрешность увеличилась до 32%. Неустойчивость алгоритма разделения энергии относительно внешних шумов отмечалась и в [3].

По всей видимости, эти факторы являются причиной того, что данные о частотных модуляциях, полученные разными авторами с помощью разных методов, зачастую плохо согласуются друг с другом. Например, согласно [8], частотные модуляции для первых двух формантных частот находятся в диапазоне 0.3–19% (первая форманта) и 4–28.5% (вторая форманта). В [14] приводятся иные диапазоны – 13–24% для первой форманты, 14–40% для второй форманты, 9–40% для третьей форманты. В той же работе, а также в работе [15], сообщается, что частотные модуляции зависят от диктора и типа гласного. Напротив, в [3] эти зависимости не были обнаружены.

Во всех известных работах по анализу частотных модуляций, кроме [8, 9], вид этих модуляций нестабилен от импульса к импульсу голосового источника.

Причины частотных модуляций при фиксированной форме речевого тракта, скорее всего, многообразны и не связаны с каким-либо единственным механизмом. Можно предположить существование следующих взаимосвязанных механизмов.

1. Параметрическое изменение резонансных частот речевого тракта вследствие изменения граничных условий при открытой голосовой щели.

2. Колебания стенок речевого тракта.

3. Взаимодействие импульсов возбуждения голосового источника с трактом, приводящее к сдвигу мгновенных частот в речевом сигнале.

4. Появление при открытой голосовой щели акустических колебаний, частота которых определяется свойствами подсвязочной области – трахеи, бронхов и легких.

Влияние граничных условий со стороны голосовой щели на резонансные частоты тракта обсуждалось в ряде работ. Различные модели взаимодействия речевого тракта и подсвязочной области подтверждают возможность возникновения частотных модуляций, хотя и в разных диапазонах. Так, в работе [16] были получены относительно малые изменения первой резонансной частоты (0.2–1%) при площади голосовой щели равной 0.027 см^2. Измеренная в прямых экспериментах площадь голосовой щели может доходить до 0.2 см^2 . В диапазоне этих величин в [17] были обнаружены существенно большие смещения частоты первого резонанса (табл. 1).

В работе [18] было показано, что при некотором соотношении импедансов речевого тракта и подсвязочной области, и с учетом переменной скорости звука в голосовой щели частота первого резонанса однородной акустической трубы увеличивается на 9.2% при открытой голосовой щели площадью 0.2 см^2. Там же было установлено, что знак частотной модуляции может смениться на обратный при определенных условиях. Это означает, что при открытой голосовой щели резонансная частота может уменьшиться вместо возрастания.

Результаты взаимодействия речевого тракта и подсвязочной области не ограничиваются наблюдаемыми амплитудно-частотными модуляциями формант. В [18] было найдено, что при раскрытии голосовой щели создаются условия для развития дополнительных резонансов и антирезонансов. В результате этого на интервале открытой голосовой щели в речевом сигнале появляются спектральные компоненты, которые отсутствуют при закрытой голосовой щели. Согласно [19, 20], взаимодействие резонансов речевого тракта и подсвязочной области может привести к скачкооб-

разным изменениям формантных треков, иногда наблюдаемым на сонограммах речевых сигналов. При этом амплитуда этих скачков может достигать 300 Гц. Поскольку на периоде основного тона форма речевого тракта меняется мало, то использование различных значений резонансных частот тракта при открытой и закрытой голосовой щели могло бы способствовать более устойчивому решению обратной задачи относительно формы речевого тракта.

Упомянутые работы по исследованию влияния граничных условий и подсвязочной области были выполнены в основном в середине 80-х годов XX века на сравнительно простых математических моделях акустических процессов речеобразования, и требуют более детального рассмотрения. Вместе с тем, нам неизвестны работы по теоретическому анализу других механизмов возникновения и экспериментальной оценке количественных значений частотных модуляций.

Поскольку ожидаемые изменения формантных частот могут быть относительно невелики, то предъявляются довольно жесткие требования к точности определения самих формантных частот. В ряде работ используются квазианалитические методы их оценки, где принимается представление речевого сигнала $f(t)$ в форме суперпозиции откликов $\psi_i(t)$ резонансов речевого тракта на возбуждение:

$$f(t) = \sum_{i=1}^{N} \psi_i(t),$$

где N – число резонансов в заданном частотном диапазоне. Вид функций $\psi_i(t)$ выбирается в зависимости от конкретного метода. В методах линейного предсказания предполагается, что функции $\psi_i(t)$ представляют собой затухающие гармонические колебания с постоянной частотой на каждом периоде основного тона. В работе [1] эти функции описываются в более общем виде:

$$\psi_i(t) = a_i(t)\cos\left(2\pi[\Omega_i t + \int_0^t q_i(\tau)d\tau] + \theta_i\right),$$

где Ω_i – медленно меняющаяся частота i-го резонанса, θ_i – начальная фаза сигнала, а q_i – частотная модуляция. Такое представление лежит в основе метода разделении энергии.

В принципе, желателен такой метод оценки формантных частот, который не опирался бы на какую-либо математическую модель сигнала. Простейший способ прямого вычисления мгновенной частоты резонанса состоит в следующем. Звуковой сигнал регистрируется в виде функции $f(t) = P(t) - P_0$, описывающей изменение звукового давления $P(t)$ относительно исходного среднего давления P_0. Находя нули τ_n, $n = 1, 2, ...$, функции $f(t)$, можно дать предварительные оценки формантных частот в виде $F_n = 1/(2(\tau_{n+1} - \tau_n))$. Сортируя эти числа по частотным диапазонам, в которых предположительно находится только одна резонансная частота, и, анализируя их распределение в таких диапазонах, можно получить надежную и устойчивую динамическую оценку для формант. С подобных методов начинались работы по автоматическому распознаванию речи (см. например, [21]). В дальнейшем этот подход не получил распространения из-за того, что необходимо автоматическое определение частотной области для оцениваемого резонанса. Однако в задаче оценки частотных модуляций указание этих областей возможно вручную и метод эффективно применим. Сравнительный анализ этого "метода нулей сигнала" с другими дан ниже в п. 4. Мы будем использовать метод нулей при исследовании механизмов возникновения частотных модуляций в разных условиях и для верификации моделей процессов, порождающих эти модуляции.

Таблица 2. Диапазоны формантных частот гласных русского языка для мужских голосов

Гласный	F_1, Гц	F_2, Гц	F_3, Гц
А	450–850	950–1500	1900–2950
Э	320–530	1450–2250	2000–2950
О	300–750	600–1400	1800–3200
И	200–550	1650–2750	2250–3500
Ы	210–500	1650–2600	2150–3100
*Е**	250–570	1450–2550	2150–3350
*Я**	330–750	1350–2200	2000–3100

2. ЭКСПЕРИМЕНТЫ ПО ОЦЕНКЕ МОДУЛЯЦИЙ

Была проведена серия экспериментов, направленных на выявление модуляций формант для реального речевого сигнала. Анализ выполнялся для речевых сегментов ударного гласного в составе числительных русского языка для слогов, содержащих переходы от одного гласного к другому. В предыдущих исследованиях были установлены примерные диапазоны значений первых трех формант для каждого такого гласного (см. табл. 2, 3). Гласный /*У*/ не представлен в этих таблицах, так как он отсутствует в используемой базе данных для числительных русского языка от 0 до 9. Звуки, отмеченные звездочкой, находятся в позиции между мягкими согласными.

Речевые сигналы для известных гласных пропускались через систему перекрывающихся КИХ-фильтров с окном Хемминга, частотный

Таблица 3. Диапазоны формантных частот гласных русского языка для женских голосов

Гласный	F_1, Гц	F_2, Гц	F_3, Гц
А	550–1000	1100–1650	1950–3100
Э	350–600	1800–2600	2350–3350
О	320–850	600–1550	1800–3300
И	220–620	1850–3100	2550–3600
Ы	250–580	1900–2950	2300–3600
*Е**	300–650	2000–2950	2650–3650
*Я**	400–900	1800–2650	2300–3500

диапазон каждого из которых устанавливался в соответствии с табл. 2 или 3 в зависимости от мужского или женского голоса. В каждом частотном диапазоне вычислялись интервалы между соседними моментами обращения отфильтрованного сигнала в нуль, и на основе этого проводилась оценка формантной частоты. Привязка к интервалу закрытой голосовой щели выполнялась путем вычисления огибающей сигнала с помощью преобразования Гильберта:

$$H_i(t) = \sqrt{f_i^2(t) + \tilde{f}_i^2(t)}, \qquad (1)$$

где $f_i(t)$ – сигнал в i-й частотной полосе, а $\tilde{f}_i(t)$ – тот же сигнал, сдвинутый по фазе на $\pi/2$. Интервал времени, на котором значения такой огибающей близки к максимуму, примерно соответствует времени действия источника голосового возбуждения акустических колебаний в речевом тракте. Аналогично, интервал вблизи минимума огибающей сопоставляется режиму закрытой голосовой щели и свободным колебаниям с собственными частотами тракта.

Приведем некоторые результаты обработки данных методом нулей сигнала, упомянутым выше (его детали описаны в [22]). На рис. 1 показаны вычисленные мгновенные значения первых трех резонансных частот для слога /*ИА*/, произнесенного в условиях низких внешних шумов. На рис. 1 отчетливо выражены модуляции формантных частот. На рис. 2 те же модуляции показаны в другом масштабе времени. Нижняя кривая изображает огибающую речевого сигнала в нижней полосе частот, вычисленную по формуле (1), и представленную в некотором условном масштабе. Как видно из рис. 2, в моменты пиков огибающей частоты первой и второй формант максимальны ($F_1 \approx 660$ Гц, $F_2 \approx 1870$ Гц). В области минимума огибающей формантные частоты примерно равны $F_1 \approx 550$ Гц, $F_2 \approx 1340$ Гц. Отсюда получим индекс частотных модуляций (т.е. относительное отклонение максимальной частоты от минимальной) для первой форманты – примерно 20%, а для второй форманты – около 40%.

Аналогичные отклонения были найдены и на стационарном участке сегмента /*И*/ в этом слоге. Речевой сигнал, синтезированный с помощью полученных таким путем формантных частот (на интервале закрытой голосовой щели) для измеренной на реальном сигнале частоты основного

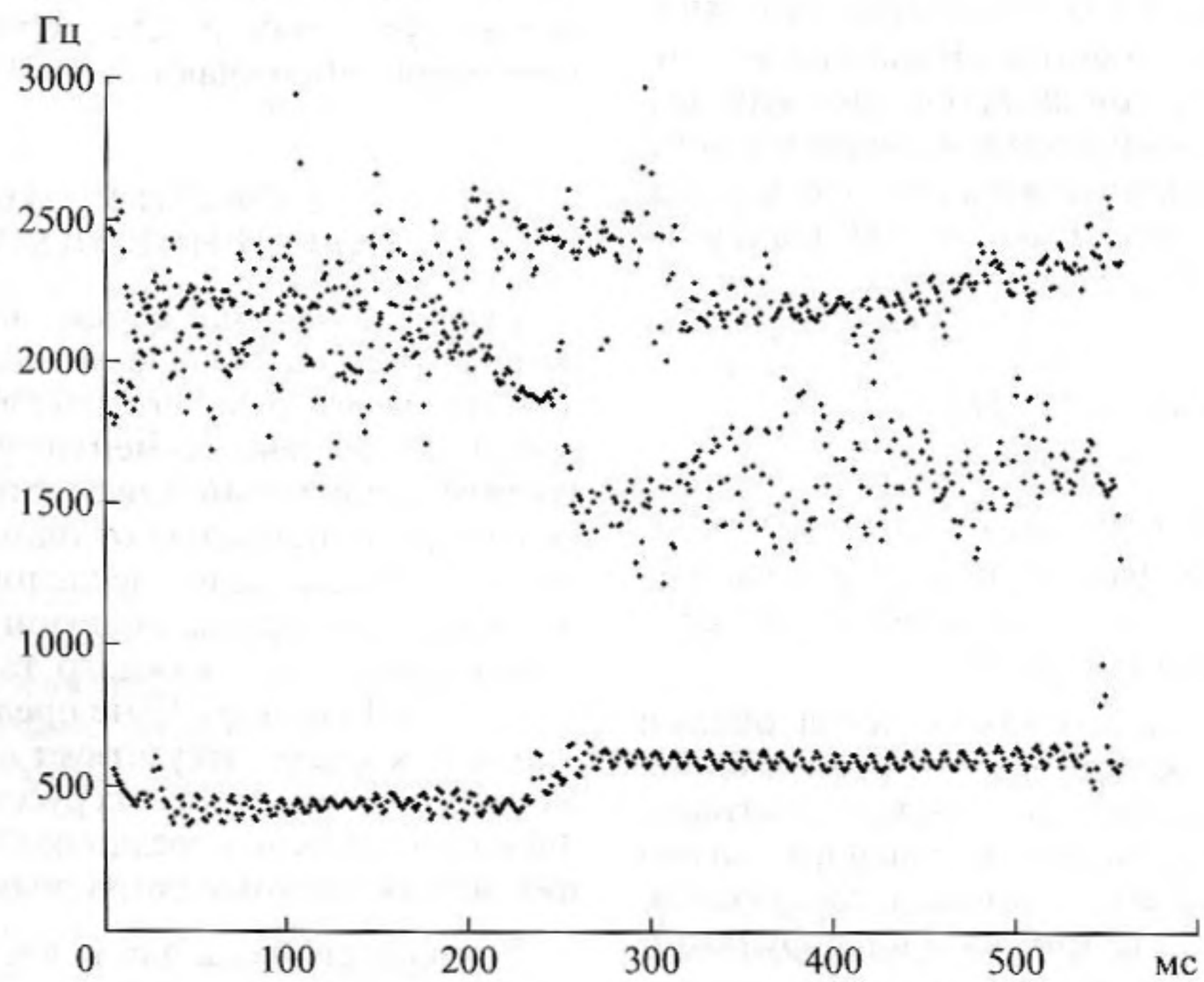

Рис. 1. Модуляции формантных частот в слоге /*ИА*/. Мужской голос.

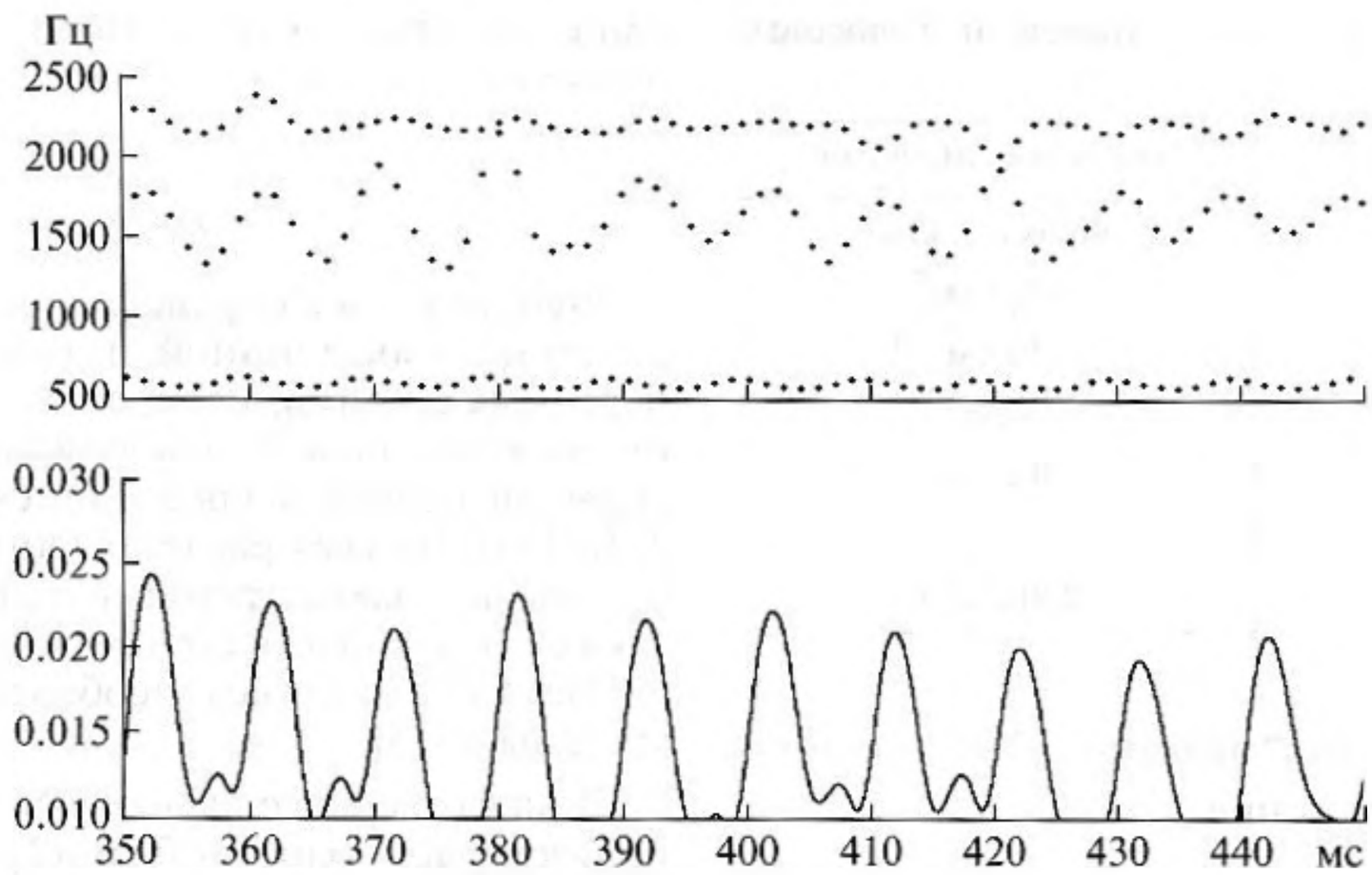

Рис. 2. Модуляции формантных частот в слоге /*ИА*/ на сегменте гласного /*А*/ (интервал 350 – 450 мс на рис. 1) (вверху). Огибающая энергии в нижней полосе частот (внизу). Частота основного тона $F_0 \approx 100$ Гц.

тона, оказался персептивно близок к исходному сигналу.

На первый взгляд этот пример показывает, что частотные модуляции, наблюдаемые в реальном речевом сигнале, соответствуют ожидаемым диапазонам значений, вычисленным по моделям параметрического воздействия переменных граничных условий на голосовой щели, использованным в работах [17, 18], и согласуются с ранее опубликованными оценками [8, 9]. Однако более детальный анализ причин явления, приведенный ниже, приводит к утверждению, что стабильная оценка модуляций определяется и другими факторами, причем она возможна далеко не всегда.

3. МОДЕЛИ ПРОЦЕССОВ, ПОРОЖДАЮЩИХ МОДУЛЯЦИЮ

Оценим некоторые факторы, которые могут порождать частотные модуляции резонансов речевого тракта.

3.1. Переменные граничные условия на голосовой щели

Простая математическая модель, объясняющая модуляции резонансных частот речевого тракта вследствие изменений граничного условия на голосовой щели, основана на хорошо известном уравнении Вебстера, которое описывает колебания в речевом тракте с медленно меняющейся геометрией (см., например, [18]). В такой постановке задачи частоты собственных колебаний речевого тракта определяются стационарным (для определенных интервалов времени) профилем площадей его поперечного сечения $S(x,t) \approx S(x)$, $0 \le x \le L$. Здесь x – координата вдоль средней линии тракта, а L – его длина. Математически эти частоты могут быть найдены путем решения соответствующей задачи на собственные значения λ_n и собственные функции $\varphi_n(x)$ задачи Штурма – Лиувилля для уравнения Вебстера:

$$\frac{1}{S(x)}\frac{\partial}{\partial x}\left(S(x)\frac{\partial \varphi_n}{\partial x}\right) + \lambda_n \varphi_n = 0,$$
$$\frac{\partial \varphi_n}{\partial x}(0) - A\varphi_n(0) = 0, \quad \frac{\partial \varphi_n}{\partial x}(L) + B\varphi_n(L) = 0. \tag{2}$$

Для простоты анализа предположим, что излучение акустических колебаний из речевого тракта в подсвязочную область примерно соответствует излучению в свободное пространство. Тогда величины A, B в краевых условиях задачи (2) задаются равенствами

$$A = \frac{3\pi^2}{8S(0)}\sqrt{\frac{S_{\text{glot}}}{\pi}}, \quad B = \frac{3\pi^2}{8S(L)}\sqrt{\frac{S_{\text{lips}}}{\pi}},$$

где S_{glot} – площадь голосовой щели, а S_{lips} – площадь тракта у губ [18].

Резонансные частоты тракта (форманты) вычисляются по собственным числам λ_n: $F_n = c_0\sqrt{\lambda_n}$, где c_0 – скорость звука в тракте. Качественно можно предсказать поведение формант при открытии голосовой щели в квазистационарном

Таблица 4. Параметры легких, трахеи и голосовых складок

Параметр	Числовое значение
R_l	40 акуст. Ом
V_l	3000 см^3
l_{tr}	14 см
S_{tr}	3 см^2
h_g	0.5 см
l_g	1.5 см
U_g	200 см^3/с

приближении, т.е. при "медленных" изменениях величины S_{glot}. Выражение

$$\lambda_n \int_0^L S(x)\varphi_n^2(x)dx = B\varphi_n^2(L)S(L) +$$

$$+ A\varphi_n^2(0)S(0) + \int_0^L S(x)\left(\frac{\partial\varphi_n(x)}{\partial x}\right)^2 dx,$$

которое получается при решении задачи Штурма–Лиувилля (2), показывает увеличение числа λ_n, т.е. формантной частоты, при увеличении числа A (при раскрытии голосовой щели). Задача (2) была численно решена для известных профилей площадей $S(x)$ основных гласных (см. [23]) при различных площадях раскрытой голосовой щели S_{glot}. Расчеты подтверждают, что увеличение мгновенных резонансных частот происходит синхронно с открытием голосовой щели. Полученный диапазон вариаций собственных частот задачи (2) в результате изменении площади голосовой щели совпадает с результатами [18].

3.2. Влияние подсвязочной области

Более сложная модель акустики речеобразования, учитывающая импеданс подсвязочной области по [18], состоит в использовании схемы длинной линии, в которой граничные условия могут зависеть от частоты. Пусть T – передаточная функция речевого тракта, вычисленная в предположении бесконечного акустического импеданса голосовой щели, Z – входной акустический импеданс в речевой тракт со стороны голосовой щели, Z_{sub} – входной акустический импеданс в трахею со стороны голосовой щели, Z_g – акустический импеданс голосовой щели. Тогда передаточная функция T_{tr} речевого тракта с учетом конечного импеданса голосовой щели и наличия подсвязочной области определяется как [20]:

$$T_{tr} = T\frac{Z_g}{Z + Z_g + Z_{sub}}.$$

Функции T и Z определялись с помощью обобщенной схемы длинной линии по площадям поперечных сечений, измеренным с помощью магнитно-резонансной томографии речевого тракта реального диктора для английских гласных /*A*, *E*, *I*, *U*/ [24]. Во всех расчетах учитывалось наличие потерь на вязкое трение и теплопроводность, а также податливость стенок речевого тракта. Числовые значения всех необходимых параметров указаны в [25].

В качестве модели подсвязочной области использовалась полость объема V_l, аппроксимирующая легкие, которая сочленялась с однородной трубой с потерями и податливыми стенками (трахея). Импеданс полости определялся как:

$$Z_l = R_l + \rho_0 c_0^2/(j\omega V_l).$$

Здесь R_l – активное сопротивление воздушному потоку в легких, ρ_0 – плотность воздуха, $j = \sqrt{-1}$, ω – круговая частота (рад/с). Зная Z_l и характеристики трахеи (ее длину l_{tr} и площадь поперечного сечения S_{tr}, а также параметры импеданса ее стенок), можно определить входной акустический импеданс Z_{sub} подсвязочных областей.

Акустический импеданс Z_g голосовой щели вычислялся по формуле, приведенной в [18]:

$$Z_g = \frac{12\mu h_g}{l_g d_g^3} + \frac{\rho_0 U_g}{S_g^2} + j\omega\frac{\rho_0 h_g}{S_g}.$$

Здесь μ – коэффициент вязкости воздуха, h_g – толщина голосовых складок, l_g – длина голосовых складок, d_g – ширина голосовой щели, U_g – среднее за период значение объемной скорости, протекающей через голосовую щель, S_g – среднее за период значение площади голосовой щели. Предполагалось, что голосовая щель с достаточной степенью точности аппроксимируется прямоугольником, площадь которого может быть вычислена как $S_g = d_g l_g$. Числовые значения параметров легких, трахеи и голосовой щели указаны в табл. 4.

Все вычисления проводились для 5-ти значений площади голосовой щели S_g: 0 см^2 (случай бесконечного импеданса голосовой щели), 0.04 см^2, 0.08 см^2, 0.12 см^2, 0.2 см^2 (случай максимально открытой голосовой щели). На рис. 3 представлены амплитудно-частотные характеристики четырех гласных звуков для указанных значений площади голосовой щели. Видно, что по мере раскрытия голосовой щели появляются дополнительные ре-

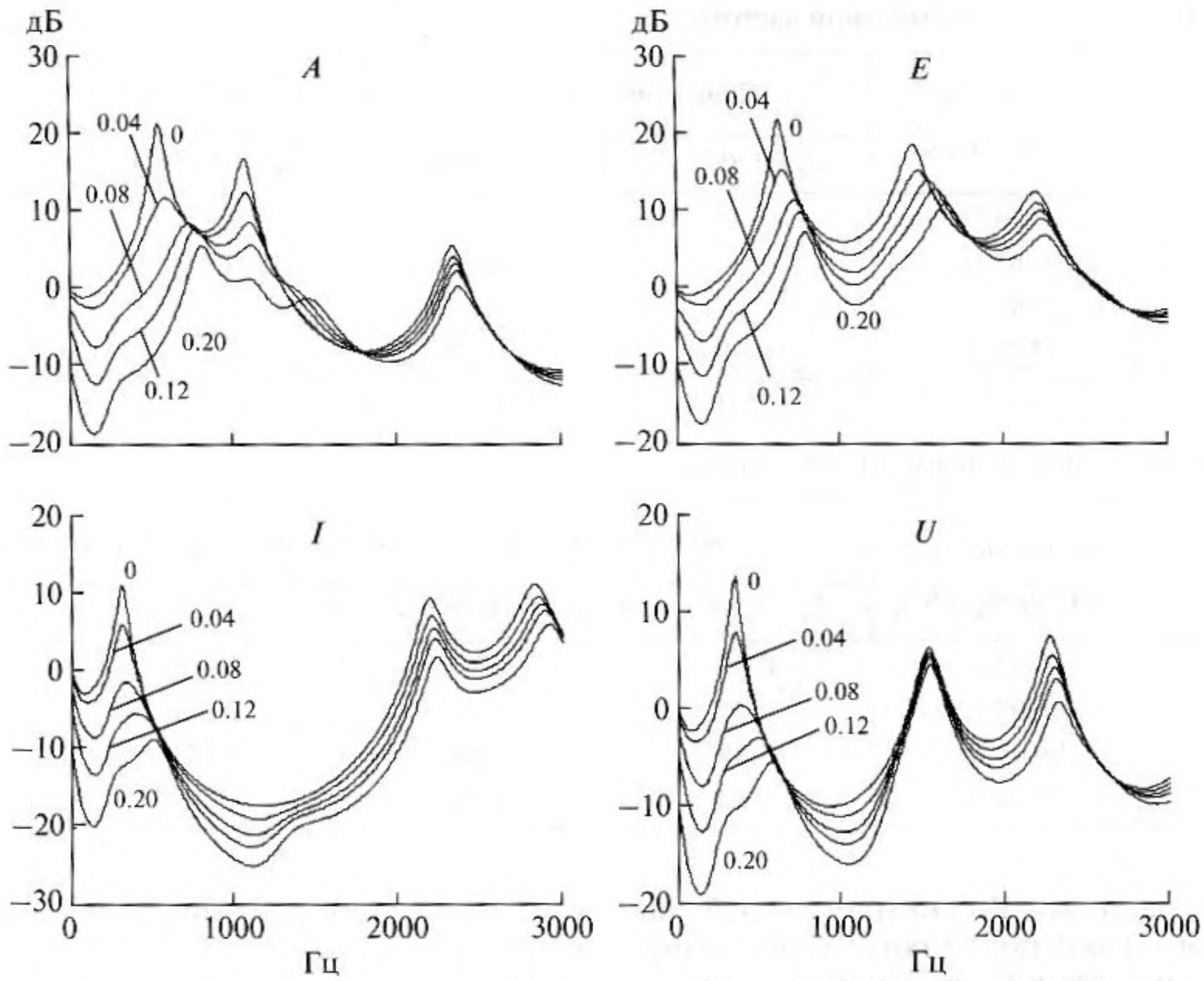

Рис. 3. Огибающие амплитудно-частотных характеристик для четырех гласных звуков и пяти значений площади голосовой щели.

зонансы и антирезонансы, а основные резонансы сдвигаются по частоте.

Частоты основных резонансов определялись по пикам соответствующих амплитудно-частотных характеристик. Девиации F_1, F_2, F_3 (в %) по отношению к частотам $F_1^{(0)}$, $F_2^{(0)}$, $F_3^{(0)}$ для сомкнутых голосовых складок указаны в табл. 5, 6 и 7.

При больших значениях площади голосовой щели речевой тракт и подсвязочная область представляют собой единую акустическую систему. Поэтому в некоторых случаях было невозможно решить, какой пик соответствует резонансу речевого тракта, а какой – подсвязочной области (рис. 3, гласные /*I*, *U*/, $S_g = 0.2$ см^2, диапазон ниже 1 кГц). В этих случаях за основной резонанс формально принимался пик, лежащий на более высокой частоте, что приводило к значительным (>60%) девиациям по F_1 для этих гласных. Возможно, что в таких случаях вообще нельзя говорить о девиациях основных формант.

3.3. Колебания стенок в сужениях тракта

Изменение резонансных частот речевого тракта могло бы возникнуть в результате колебаний стенок тракта в его сужениях, поскольку именно малые площади тракта сильнее всего влияют на

Таблица 5. Девиация первой формантной частоты

Гласный	Значение $F_1^{(0)}$ для $S_g = 0$ см^2	Девиация F_1 в % для разных значений S_g относительно $F_1^{(0)}$			
		$S_g = 0.04$ см^2	$S_g = 0.08$ см^2	$S_g = 0.12$ см^2	$S_g = 0.2$ см^2
/*A*/	560 Гц	8.5%	30.7%	39.3%	44.4%
/*E*/	634 Гц	4.5%	15.0%	19.6%	25.6%
/*I*/	310 Гц	0.1%	6.2%	40.0%	61.5%
/*U*/	348 Гц	2.7%	10.9%	41.1%	65.8%

Таблица 6. Девиация второй формантной частоты

Гласный	Значение $F_2^{(0)}$ для $S_g = 0$ см^2	Девиация F_2 в % для разных значений S_g относительно $F_2^{(0)}$			
		$S_g = 0.04$ см^2	$S_g = 0.08$ см^2	$S_g = 0.12$ см^2	$S_g = 0.2$ см^2
/A/	1074 Гц	1.8%	3.6%	4.4%	4.4%
/E/	1465 Гц	1.3%	5.9%	9.8%	13.0%
/I/	2190 Гц	0.9%	1.7%	1.7%	2.2%
/U/	1532 Гц	0.6%	0.6%	0.6%	0.6%

Таблица 7. Девиация третьей формантной частоты

Гласный	Значение $F_3^{(0)}$ для $S_g = 0$ см^2	Девиация F_3 в % для разных значений S_g относительно $F_3^{(0)}$			
		$S_g = 0.04$ см^2	$S_g = 0.08$ см^2	$S_g = 0.12$ см^2	$S_g = 0.2$ см^2
/A/	2343 Гц	0.4%	0.8%	1.2%	1.6%
/E/	2209 Гц	0.9%	1.7%	2.2%	2.6%
/I/	2845 Гц	0.8%	1.6%	2.0%	2.8%
/U/	2276 Гц	0.8%	1.3%	1.7%	2.1%

резонансные частоты. Рассмотрим колебания стенок речевого тракта под влиянием избыточного давления, создаваемого воздушным потоком, который проходит через голосовую щель синхронно с колебаниями голосовых складок. Основное влияние на резонансные частоты тракта эти колебания стенок оказывают в области наибольшего его сужения. Толщина стенок тракта значительно меньше длин волн акустических колебаний до частот 3 кГц. Поэтому в области наибольшего сужения стенки могут рассматриваться как система с сосредоточенными параметрами, и их колебания на единицу длины тракта описываются обыкновенным дифференциальным уравнением второго порядка

$$m_w y'' + r_w y' + c_w y = P_{vt}, \qquad (3)$$

где m_w, r_w, c_w – погонные масса, вязкое и упругое сопротивление, P_{vt} – переменное давление, синхронное с колебаниями голосовых складок. Значения погонной массы вычисляется из равенства $m_w = \rho_w h_w$, где $\rho_w \approx 1.1$ г/см^3 – плотность тканей, 0.5 см $\le h_w \le 2$ см, откуда 0.55 г/см^2 $\le$ $\le m_w \le 2.2$ г/см^2. Упругое сопротивление $c_w = E_w / h_w$ выражается через модуль упругости тканей E_w, который варьируется в широких пределах. Это приводит к оценке 10^4г/см^{2}с^2 $\le$ $\le c_w \le 10^5$г/см^{2}с^2. Вязкое сопротивление находится в диапазоне 800 г/см^{2}с $\le r_w \le 1100$ г/см^{2}с. Переменное давление в речевом тракте описывается системой нелинейных дифференциальных уравнений [23].

Решение дифференциального уравнения (3) для различных параметров m_w, r_w, c_w в указанных диапазонах показало, что амплитуда колебаний стенок речевого тракта находится в интервале $y_{\max} = 10^{-4} - 4\times10^{-3}$ см, что значительно меньше расстояния между поверхностями речевого тракта для гласных. В результате наибольший коэффициент частотной модуляции оказался порядка 0.25%. Аналогично, влияние потерь на колебание стенок составило величину порядка ±0.35%.

Таким образом, приближенный анализ роли колебаний стенок в модуляции резонансных частот указывает на их пренебрежимо малое влияние по сравнению с возможной погрешностью измерения.

4. СРАВНЕНИЕ МЕТОДОВ АНАЛИЗА ФОРМАНТ И УСТОЙЧИВОСТЬ ОЦЕНОК МОДУЛЯЦИЙ

Из сказанного выше следует, что возможность появления заметных изменений резонансных частот на интервале открытой голосовой щели может быть объяснена механизмами изменения граничных условий при открытии голосовой щели и влиянием подсвязочной области. В целом, это подтверждают сообщения в различных публикациях о наблюдении частотных модуляций, синхронных с импульсами источника голосового возбуждения. Вместе с тем, вид и знак наблюдаемых модуляций подвержен еще и значительным

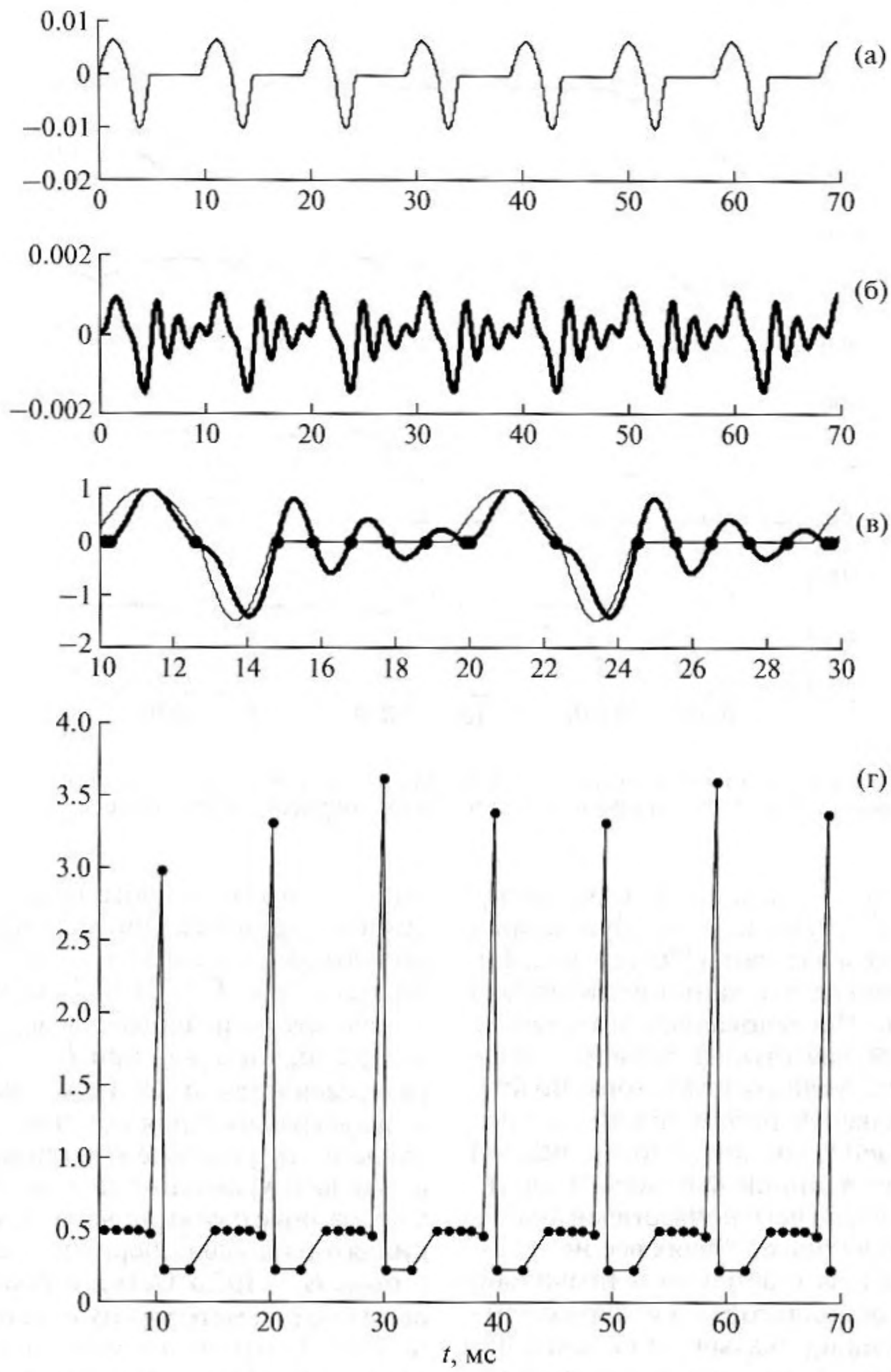

Рис. 4. Оценки частоты осциллятора методом анализа интервалов между нулями сигнала: а) голосовой источник; б) реакция осциллятора на этот источник; в) нормированные кривые а), б) и нули сигнала; г) временной трек форманты, полученный по методу нулей.

вариациям в зависимости от некоторых иных факторов. Например, в работах [8, 9] было обнаружено, что амплитуда и фаза частотных модуляций относительно пика огибающей энергии речевого сигнала на стационарном участке одного и того же гласного звука могут существенно меняться. Это вызывает сомнение, например, в том, что вычисленные оценки частотных модуляций порождаются только изменением граничных условий на периоде основного тона. Необходимо исследовать влияние различных методов анализа речевого сигнала на оценки формантных частот. С этой целью был проведен ряд экспериментов. Исследовались следующие методы оценки частотных модуляций с различными фильтрами: метод нулей сигнала, линейное предсказание второго порядка на коротком интервале (5 отсчетов сигнала) [5], метод преобразования Гильберта [2] и алгоритм разделения энергии [1, 2].

Исходными данными для применения этих методов служили модели речевого сигнала, в ко-

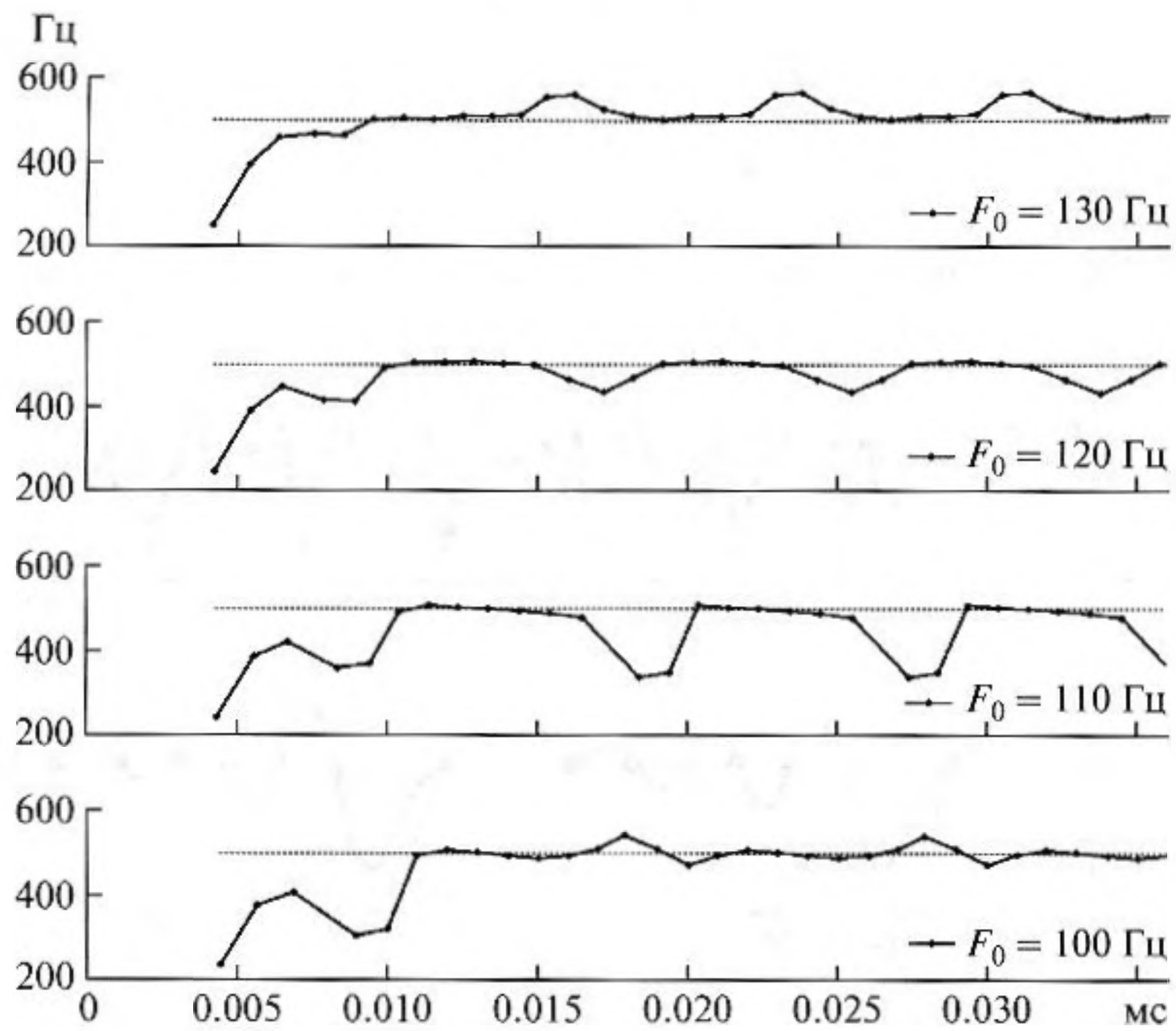

Рис. 5. Модуляции оценки частоты резонанса $F = 500$ Гц для различной частоты основного тона F_0. Использован фильтр с окном Хемминга. Пунктиром указана частота 500 Гц гармонического осциллятора.

торых использовался одиночный гармонический осциллятор с затуханием, возбуждаемый голосовым источником из работы [26] (см. рис. 4а). Частота и коэффициент затухания осциллятора не модулировались. Исследовались два режима оценки мгновенной частоты. В первом случае анализу подвергался отклик осциллятора. Во втором случае отклик сначала пропускался через полосовой фильтр (с центральной частотой, равной частоте осциллятора, и шириной полосы, равной, примерно, 400 Гц), после чего подвергался анализу. Выяснилось, что в обоих случаях все исследованные методы анализа стабильно и правильно оценивали частоту осциллятора на интервале отсутствия возбуждения и показывали наличие паразитных частотных модуляций на интервале возбуждения. Амплитуда и фаза этих модуляций существенно зависела как от соотношения частоты основного тона голосового возбуждения и собственной частоты осциллятора, так и от метода анализа и типа фильтра. Проиллюстрируем это на примере метода нулей сигнала, как наиболее устойчивого метода оценки мгновенной частоты [22].

Без использования фильтрации метод нулей сигнала показывает определенную неустойчивость по отношению к изменениям частоты основного тона. Вид и размах частотных модуляций, которые он предсказывает при фиксированной частоте резонатора, могут значительно изменяться во времени при очень малых изменениях частоты основного тона, заимствованных из реально произнесенного гласного. Так, для осциллятора с частотой $F = 500$ Гц при частоте основного тона $F_0 = 102.57$ Гц оценка частоты осциллятора на интервале возбуждения падает ниже 150 Гц, тогда как при $F_0 = 102.5$ Гц эта оценка резко меняется от 200 Гц до 3500 Гц. На рис. 4 показаны графики: для источника возбуждения (а), численного решение обыкновенного дифференциального уравнения для осциллятора (б), нормированные функции возбуждения и отклика осциллятора для двух периодов основного тона с частотой $F_0 = 102.5$ Гц (в), а также оценки частоты осциллятора методом нулей сигнала (г). Точками на рис. 4в отмечены моменты перехода сигнала через нуль. Из рис. 4в, г видно, что резкие колебания оценки форманты связаны с близкими нулями сигнала, возникающими на интервале возбуждения.

Рассмотрим теперь результаты оценки мгновенной частоты методом нулей по фильтрованному сигналу. Известно, что в зависимости от вида фильтра и крутизны его "склонов" фильтруемый сигнал искажается по-разному. Мы исследовали фильтры Баттерворта 10-го и 12-го порядка, фильтр Габора и КИХ-фильтр с окном Хемминга. Выяснилось, что помимо искажений, которые каждый фильтр вносит во временную форму сигнала, фильтр Габора также дает систематическую погрешность при оценке формантной частоты. Эта ошибка зависит от отношения частоты воз-

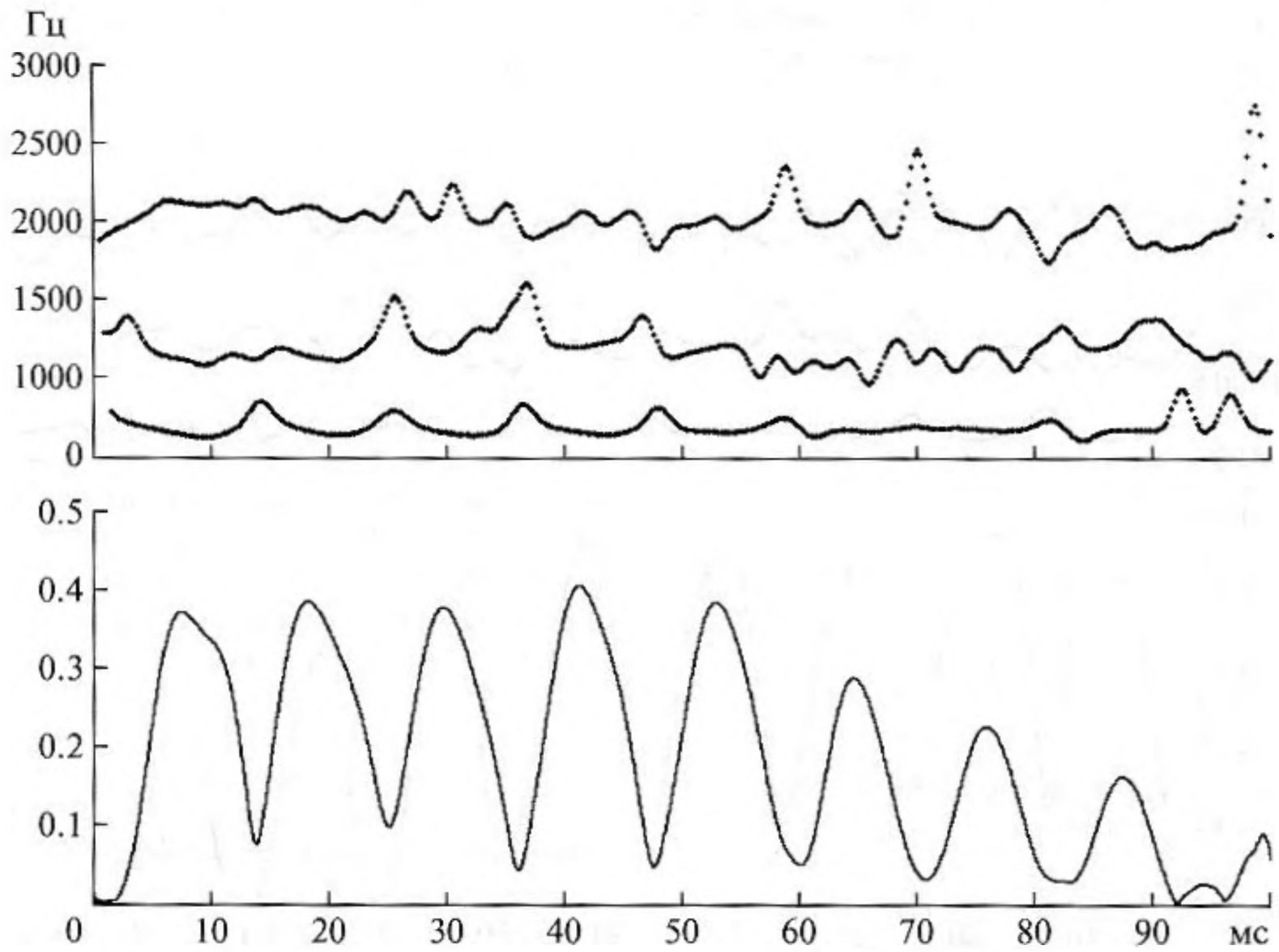

Рис. 6. Модуляции в оригинальном гласном /*A*/, кардиоидный микрофон на груди диктора. Внизу – огибающая энергии в области первой форманты (в долях от общей энергии сигнала).

буждения и частоты резонатора, а также от сдвига частоты резонатора относительно центральной частоты фильтра. Фильтр с окном Хемминга обеспечивает компромисс между необходимостью разделения соседних формантных частот и степенью искажения сигнала во временной области. Как и в первом случае (анализ без использования фильтрации), оценка мгновенной частоты методом нулей для всех типов фильтров существенно зависит от соотношения частоты возбуждения и собственной частоты гармонического осциллятора. Пример вариабельности оценок модуляций для нескольких импульсов возбуждения с разной частотой F_0 основного тона показан на рис. 5.

Результаты приведенного модельного эксперимента объясняют изменение вида частотных модуляций оценок формант на одном и том же стационарном гласном с переменной частотой основного тона, обнаруженное в [8, 9].

Аналогичное поведение оценок резонансных частот наблюдаются и при анализе другими методами. Рассмотрим, например, метод линейного предсказания для оценки формант. Сигнал на интервале возбуждения осциллятора с переменной частотой можно приближенно описать в спектральной области произведением меняющихся во времени комплексных спектров источника возбуждения и осциллятора. Реакция метода линейного предсказания на такое искажение спектра хорошо известна: появляется так называемый сигнал-остаток, а оценка испытывает резкие скачки [27].

Таким образом, даже несмотря на фильтрацию, при анализе речевого сигнала могут появляться "паразитные" частотные модуляции, которые определяются исключительно методом этого анализа и не имеют отношения к физически обусловленным модуляциям. Можно ожидать, что подобные паразитные модуляции будут искажать или маскировать истинные вариации резонансных частот тракта на интервале открытой голосовой щели.

На вид оценок модуляций в речевом сигнале влияет не только метод анализа, но и тип приемника звука. Модуляции в одном и том же гласном, записанном параллельно через два микрофона разного типа, могут существенно отличаться при довольно близких значениях оценок формантных частот на интервале закрытой голосовой щели [22]. Очевидно, что это отличие связано не с реальными характеристиками процесса речеобразования, а только с искажениями речевого сигнала в микрофоне. Реверберация помещения также сказывается на параметрах речевого сигнала, если микрофон удален от губ диктора.

На рис. 6 и 7 показано различие в модуляциях речевого сигнала, параллельно записанного через микрофоны разного типа. Видно, что модуляции третьей форманты в сигналах от разных микрофонов заметно отличаются и эти модуляции не

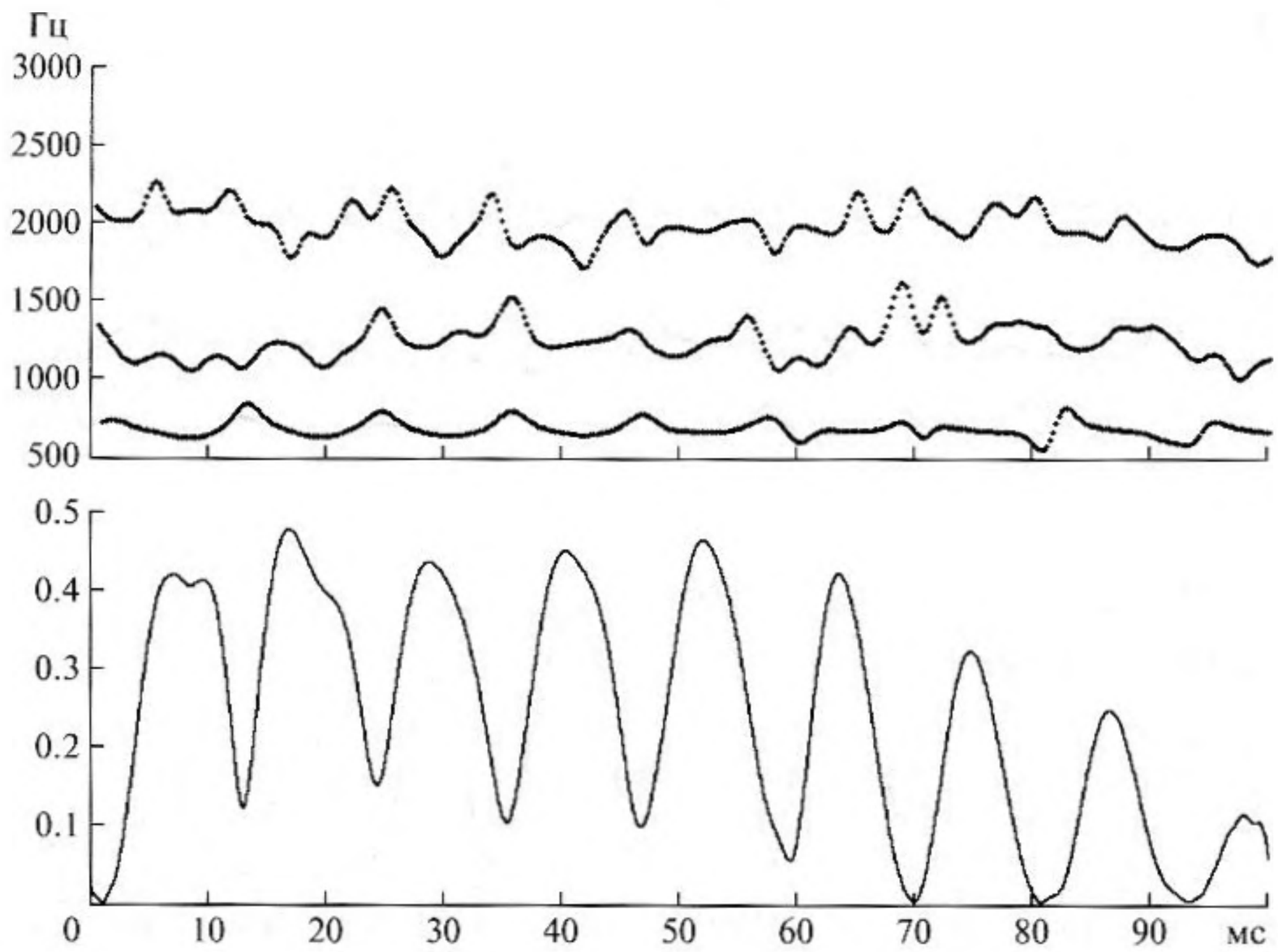

Рис. 7. Модуляции в оригинальном гласном /*A*/, телефонная трубка. Внизу – огибающая энергии в области первой форманты (в долях от общей энергии сигнала).

синхронны с импульсами голосового возбуждения. Существует также заметное различие в форме модуляций второй форманты, и лишь модуляции первой форманты более или менее похожи на участке до 60 мс. Модуляции в сигнале, записанном через микрофон другого типа, также на сегменте гласного /*A*/, произнесенного другим диктором (рис. 7), демонстрируют значительно более ясную картину модуляций для всех трех формант синхронно с импульсами голосового источника.

Вид частотных модуляций, показанных на рис. 7, значительно более соответствует ожидаемым процессам, связанным с влиянием импеданса голосовой щели и подсвязочной области, как это следует из описанных выше моделей их взаимодействия с речевым трактом. Поэтому визуально часто можно определить, вызваны ли измеренные модуляции граничными условиями на нижнем конце тракта или же доминируют эффекты, связанные с частотой основного тона и типом приемника звука. К сожалению, даже в первом случае неизвестна погрешность оценки резонансных частот тракта на интервале открытой голосовой щели. Необходимы дополнительные исследования для того, чтобы установить возможность автоматического выбора именно тех модуляций, которые, скорее всего, связаны с физическими механизмами речеобразования.

5. ЗАКЛЮЧЕНИЕ

Частотные модуляции, которые наблюдаются в речевом сигнале, имеют двоякую природу. Математический анализ и компьютерное моделирование позволяют выделить физические механизмы, ответственные за модуляции резонансных частот речевого тракта синхронно с колебаниями площади голосовой щели. Колебания стенок речевого тракта в сужениях оказывают пренебрежимо малое влияние на резонансные частоты тракта, тогда как изменение граничных условий при раскрытии голосовой щели и влияние подсвязочной области может привести к весьма значительным модуляциям резонансных частот. Величина этих модуляций зависит от формы речевого тракта и частот его резонансов при закрытой голосовой щели.

Модуляции мгновенных частот в речевом сигнале могут существенно отличаться по своему размаху и виду от модуляций, вызванных только изменением граничных условий в тракте. Модуляция определяется также и влиянием формы импульсов голосового источника и периода их следования на свойства речевого сигнала. Метод оценки мгновенных частот, амплитудно-частотные искажения в акустическом канале и тип приемника звука в свою очередь влияют на вид регистрируемых модуляций.

СПИСОК ЛИТЕРАТУРЫ

1. *Maragos P., Kaiser J.F., Quatieri T.F.* Energy separation in frequency modulations with application to speech analysis // IEEE Trans. Signal Process. 1993. V. 41. P. 3024–3051.
2. *Potamianos A., Maragos P.* Speech Formant Frequency and Bandwidth Tracking Using Multiband Energy Demodulation // J. Acoust. Soc. Amer. 1996. V. 99. № 6. P. 3795–3806.
3. *Dimitriadis D., Maragos P.* Continuous energy demodulation methods and application to speech analysis // Speech Comunication. 2006. V. 48. P. 819–837.
4. *Ramalingam C.* On the Equivalence of DESA-1a and Prony's Method when the Signal is a Sinusoid // IEEE Signal Process. Letters. 1996. V. 3. № 5. P. 141–143.
5. *Fertig L., McClellan J.* Instantaneous Frequency Estimation Using Linear Prediction with Comparisions to the DESAs // IEEE Signal Process. Letters. 1996. V. 3. № 2. P. 54–56.
6. *Kumaresan R., Rao A.* Model-based Approach to Envelope and Positive-Instantaneous Frequency of Signals and its Application to Speech // J. Acoust. Soc. Amer. 1999. V. 105. № 3. P. 1912–1924.
7. *Rao A., Kumaresan R.* On decomposing speech onto modulated components // IEEE Trans. Speech, Audio Process. 2000. V. 8. № 3. P. 240–254.
8. *Сорокин В.Н., Трифоненков И.П.*. Об автокорреляционном анализе речевых сигналов // Акуст. журн. 1996. Т. 42. № 3. С. 368–374.
9. *Леонов А.С., Сорокин В.Н.* К анализу резонансных частот речевого тракта // Информационные процессы. 2007. Т. 7. № 4. С. 386–400.
10. *Kumaresan R., Wang Y.* On Representing Signals Using Only Timing Information // J. Acoust. Soc. Amer. 2001. V. 110. P. 2421–2439.
11. *Wang Y., Kumaresan R.* Real Time Decomposition of Speech into Modulated Components // J. Acoust. Soc. Amer. Express Letters. 2006. V. 119. № 6. P. EL68–EL73.
12. *Gianfelici F., Biagetti G., Crippa P., Turchetti C.* Multicomponent AM-FM representations: an asymptotically exact approach // IEEE Trans. Audio, Speech, Language Process. 2007. V. 15. № 3. P. 823–837.
13. *Vakman D.* On the Analytic Signal, the Teager-Kaiser Energy Algorithm, and Other Methods for Defining Amplitude and Frequency // IEEE Trans. Signal Process. 1996. V. 44. № 4. P. 791–797.
14. *Potamianos A., Maragos P.* Speech analysis and synthesis using AM-FM modulation model // Speech Comunication. 1999. V. 28. P. 195–209.
15. *Grimaldi M., Cummins F.* Speech Identification Using Instantaneous Frequencies // IEEE Trans. Audio, Speech, Language Process. 2008. V. 16. № 6. P. 1097–1111.
16. *Badin P., Fant G.* Notes on vocal tract computation // STL-QPSR. 1984. V. 2–3. P. 53–108.
17. *Fant G., Wakita H.* Toward a better vocal tract model // STL-QPSR. 1978. P. 9–29.
18. *Сорокин В.Н.* Теория речеобразования. М.: Радио и связь, 1985. 312 с.
19. *Stevens K.* Acoustic Phonetics // The MIT Press. 1998. P. 614.
20. *Chi X., Sonderegger M.* Subglottal coupling and its influence on vowel formants // J. Acoust. Soc. Amer. 2007. V. 122. P. 1735–1745.
21. *Цемель Г.И.* Опознавание речевых сигналов. М.: Наука, 1971. 148 с.
22. *Леонов А.С., Макаров И.С., Сорокин В.Н.* Устойчивость оценок формантных частот // Речевые технологии. 2009. № 1. С. 3–15.
23. *Сорокин В.Н.* Синтез речи. М.: Наука. 1992. 392 с.
24. *Baer T., Gore J., Gracco V., Nye P.* Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels // J. Acoust. Soc. Amer. 1991. V. 90. P. 799–828.
25. *Макаров И. С.* Аппроксимация речевого тракта коническими рупорами // Акуст. журн. 2009. Т. 55. № 2. С 256–265.
26. *Сорокин В.Н., Макаров И.С.* Определение пола диктора по голосу // Акуст. журн. 2008. Т. 54. № 4. С. 659–668.
27. *Yegnanarayana B., Veldhuis R.* Extraction of Vocal-Tract System Characteristics from Speech Signals // IEEE Trans. Speech, Audio Process. 1998. V. 6. № 4. P. 313–327.