

===== ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ТЕХНИЧЕСКИХ =====
===== И СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ СИСТЕМАХ =====

Артикуляторный ресинтез фрикативных

А.С.Леонов*, И.С.Макаров**, В.Н.Сорокин**, А.И.Цыплихин**

*Московский инженерно-физический институт, Москва, Россия

**Институт проблем передачи информации, Российская академия наук, Москва, Россия

Поступила в редколлегию 05.2004

Аннотация. Решались две обратные задачи восстановления формы речевого тракта для фрикативных согласных звуков английского языка /s, sh, f, θ, ð, h, z, zh/. В одной задаче входными данными служили только акустические параметры речевого сигнала, а в другой – вместе с акустическими параметрами использовались и траектории движений 8 точек на внутренних поверхностях речевого тракта, измеренные на микролучевом рентгенокопе. В обеих задачах погрешность оценки измеренных координат точек составляла, в среднем, меньше 3%, причем разница между решениями в пространстве артикуляторных параметров была меньше 4%, за одним исключением. Перцептивный контроль качества решения обратной задачи выполнялся путем синтеза слогов “гласный – фрикативный - гласный” артикуляторным синтезатором. Исходными данными для синтеза служили форма речевого тракта и площади его поперечного сечения, вычисленные в результате решения обратной задачи. Звучание синтезированных слогов оказалось весьма близко к оригинальным слогам.

1. Введение

Обратная задача: найти по акустическим данным речевого сигнала форму речевого тракта - для фрикативных звуков оказывается гораздо более сложной, чем для гласных. При решении такой задачи для гласных можно использовать измеренные резонансные частоты речевого тракта как исходные данные. Хотя и имеется неоднозначная зависимость формы речевого тракта от его резонансов, достаточно точные решения все же можно получить. Это удается сделать за счет использования адекватной модели речеобразования и ограничений на анатомические и артикуляторные параметры с помощью специальных алгоритмов решения обратных задач [1]. Для фрикативных звуков появляется дополнительный фактор неоднозначности, который связан с турбулентностью воздушного потока в речевом тракте и с нестабильностью спектра соответствующего шумового источника возбуждения. Трудности решения обратной задачи для фрикативных, по-видимому, являются причиной скудности научных публикаций в этой области. В мировой литературе известна только одна работа [2] по данному вопросу. В этой работе исходными данными для решения служили спектры фрикативных, усредненные на интервале стационарности. По ним определялась форма речевого тракта, по которой затем вычислялся спектр соответствующего фрикативного звука. Задача решалась вариационным методом. Критерий оптимальности состоял в минимизации работы артикуляторов, а в качестве меры сходства вычисленных и измеренных спектров использовался функционал, с точностью до постоянного множителя равный норме в пространстве L_2 . Было установлено, что вычисленный спектр был ближе к спектру истинного фрикативного, чем к спектрам других фрикативных, только в том случае, когда начальное приближение для формы тракта (в пространстве артикуляторных параметров) выбиралось из области, соответствующей искомому фрикативному. Поскольку истинная форма речевого тракта для записанных речевых сигналов не была известна, то нельзя было оценить точность решения обратной задачи на артикуляторном уровне.

С появлением возможности измерений координат некоторых точек внутри речевого тракта (с помощью микролучевого рентгенокопа или электромагнитной системы) синхронно с записью речевого сигнала появилась и возможность контроля качества решений обратной задачи для фрикативных звуков.

2. Математические модели артикуляции и акустики в прямой задаче

2.1. Модель артикуляции

Математическая модель артикуляции для фрикативных звуков не отличается от аналогичной модели для гласных. Ее подробное описание дано в [3]. Модель использует 14 управляемых артикуляторных параметров: 1) высота голосовой щели, 2-3) координаты корня языка, 4) угол поворота языка как твердого тела, 5-6) угол поворота и горизонтальное смещение точки вращения нижней челюсти, 7-11) 5 коэффициентов при собственных функциях языка, 12) высота нижней губы, 13-14) 2 коэффициента при собственных функциях ширины глотки.

Для глухих фрикативных использовалась постоянная ширина голосовой щели, равная 0.4 см^2 , а для звонких фрикативных максимальное значение площади голосовой щели при голосовом возбуждении равнялось 0.2 см^2 .

Помимо артикуляторных параметров, в модель включались следующие анатомические параметры диктора, для которого решается обратная задача:

- форма свода твердого неба,
 - ширина нижней челюсти у коренных зубов,
 - длина нижней челюсти от основания до передних зубов,
 - толщина коренных и передних зубов,
 - форма поверхности губ,
 - форма нижней челюсти в латеральной плоскости (аппроксимировалась квадратичной функцией).
- Эти параметры использовались при расчете площадей поперечного сечения тракта после нахождения его формы в среднесагиттальной плоскости. Особенности вычисления площади тракта в глотке описаны в [4].

2.2. Акустическая модель

Из теории речеобразования известно, что в диапазоне частот примерно до 4.5 кГц акустическое давление в речевом тракте описывается одномерным волновым уравнением (уравнением Вебстера):

$$\frac{\partial^2 P}{\partial t^2} = c_0^2 \frac{1}{S(x)} \frac{\partial}{\partial x} \left(S(x) \frac{\partial P}{\partial x} \right) + F(x, t), \quad 0 < x < l, t > 0 \quad (1)$$

Здесь x - пространственная координата вдоль средней линии тракта в среднесагиттальной плоскости, t - момент времени, $P(x, t)$ - искомое давление в тракте, $S(x)$ - профиль площадей поперечного сечения вдоль тракта, $F(x, t)$ - плотность распределения источников возбуждения внутри тракта, c_0 - скорость звука в тракте. Такая модель в принципе пригодна не только для гласных звуков речи, но и для фрикативных звуков, источником возбуждения которых служит шум турбулентного воздушного потока. Однако она прямо не учитывает потерь в тракте. Поскольку для гласных коэффициенты потерь значительно меньше соответствующих резонансных частот, то оказалось возможным учитывать затухания резонансных колебаний после решения уравнения (1). Это обеспечивает достаточно точное воспроизведение акустических характеристик гласных [1]. Потери для фрикативных звуков значительно выше, поэтому их необходимо учитывать непосредственно при нахождении вынужденных колебаний тракта, т.е. включить потери в тракте уже в само исходное уравнение колебаний.

Используем с этой целью известные уравнения, связывающие давление в тракте $P(x, t)$ и объемную скорость $U(x, t)$:

$$S(x) \frac{\partial P}{\partial x} + \rho \frac{\partial U}{\partial t} + r_1 U = 0, \quad \rho c_0^2 \frac{\partial U}{\partial x} + S(x) \frac{\partial P}{\partial t} + r_2 P = 0 \quad (2)$$

где ρ - плотность воздуха, а r_1, r_2 - коэффициенты потерь на вязкое трение и теплопроводность стенок тракта соответственно. Для фрикативных звуков основные потери определяются вязким трением, пропорциональным скорости колебаний воздушных частиц, тогда как потери на теплопроводность, пропорциональные давлению, невелики. Поэтому, полагая $r_2 = 0$, получим путем исключения величины $U(x, t)$ из (2) модифицированный вариант уравнения Вебстера:

$$\frac{\partial^2 P}{\partial t^2} = c_0^2 \frac{1}{S(x)} \frac{\partial}{\partial x} \left(S(x) \frac{\partial P}{\partial x} \right) - \frac{r_1}{\rho} \frac{\partial P}{\partial t}$$

С учетом источников возбуждения внутри тракта, а также вместе с краевыми и начальными условиями, получим начально-краевую задачу для модифицированного уравнения Вебстера:

$$\begin{aligned} \frac{\partial^2 P}{\partial t^2} &= c_0^2 \frac{1}{S(x)} \frac{\partial}{\partial x} \left(S(x) \frac{\partial P}{\partial x} \right) - 2\nu \frac{\partial P}{\partial t} + F(x, t), \quad 0 < x < l, t > 0 \\ \frac{\partial P}{\partial x} \Big|_{x=0} &= -\dot{q}(t); \quad \left(\frac{\partial P}{\partial x} - bP \right) \Big|_{x=l} = 0 \\ P(x, 0) &= P_0(x); \quad \frac{\partial P}{\partial t}(x, 0) = P_1(x) \end{aligned}$$

Здесь $\nu = \frac{r_1}{2\rho}$, $q(t)$ - краевой источник возбуждения (голосовой источник), $P_{0,1}(x)$ - начальные профили давления и скорости его изменения в тракте. Анализ этой задачи удобно выполнять в предположении постоянства величины ν , коэффициента b , связанного с площадью излучающего звук конца тракта, а также полагая длину тракта l постоянной на протяжении рассматриваемого промежутка времени.

Введем характерный параметр времени t_0 , равный длине интервала, на котором решается волновое уравнение. Это позволяет определить безразмерные переменные $\xi = x/l, \tau = t/t_0, k(\xi) = S(x)/l^2, \dot{Q}(\tau) = l\dot{q}(t)$ и параметры $a = c_0 t_0 / l, \delta = \nu t_0, h = b t_0$, а также соответствующие функции новых переменных (ξ, τ) :

$$p(\xi, \tau) = P(x, t), p_0(\xi) = P_0(x), p_1(\xi) = t_0 P_1(x), f(\xi, \tau) = F(x, t) t_0^2.$$

Таким образом получаем обезразмеренную по аргументам и коэффициентам начально-краевую задачу:

$$\begin{aligned} \frac{\partial^2 p}{\partial \tau^2} &= a^2 \frac{1}{k(\xi)} \frac{\partial}{\partial \xi} \left(k(\xi) \frac{\partial p}{\partial \xi} \right) - 2\delta \frac{\partial p}{\partial \tau} + f(\xi, \tau), \quad 0 < \xi < 1, \tau > 0 \\ \frac{\partial p}{\partial \xi} \Big|_{\xi=0} &= -\dot{Q}(\tau); \quad \left(\frac{\partial p}{\partial \xi} - hp \right) \Big|_{\xi=1} = 0 \\ p(\xi, 0) &= p_0(\xi); \quad \frac{\partial p}{\partial \tau}(\xi, 0) = p_1(\xi) \end{aligned} \tag{3}$$

Решение линейной задачи (3) представимо в виде суммы: $p(\xi, \tau) = u(\xi, \tau) + v(\xi, \tau) + w(\xi, \tau)$, где слагаемые справа являются решениями следующих задач:

а) задача с начальным возбуждением тракта

$$\begin{aligned} \frac{\partial^2 u}{\partial \tau^2} &= a^2 \frac{1}{k(\xi)} \frac{\partial}{\partial \xi} \left(k(\xi) \frac{\partial u}{\partial \xi} \right) - 2\delta \frac{\partial u}{\partial \tau}, \quad 0 < \xi < 1, \tau > 0 \\ \frac{\partial u}{\partial \xi} \Big|_{\xi=0} &= 0; \quad \left(\frac{\partial u}{\partial \xi} - hu \right) \Big|_{\xi=1} = 0 \\ u(\xi, 0) &= p_0(\xi); \quad \frac{\partial u}{\partial \tau}(\xi, 0) = p_1(\xi) \end{aligned}$$

б) задача с краевым возбуждением тракта голосовым источником

$$\begin{aligned} \frac{\partial^2 v}{\partial \tau^2} &= a^2 \frac{1}{k(\xi)} \frac{\partial}{\partial \xi} \left(k(\xi) \frac{\partial v}{\partial \xi} \right) - 2\delta \frac{\partial v}{\partial \tau}, \quad 0 < \xi < 1, \tau > 0 \\ \left. \frac{\partial v}{\partial \xi} \right|_{\xi=0} &= -\dot{Q}(\tau); \quad \left. \left(\frac{\partial v}{\partial \xi} - hv \right) \right|_{\xi=1} = 0 \\ v(\xi, 0) &= 0; \quad \frac{\partial v}{\partial \xi}(\xi, 0) = 0 \end{aligned}$$

в) задача о возбуждении тракта распределенным (фрикативным) источником

$$\begin{aligned} \frac{\partial^2 w}{\partial \tau^2} &= a^2 \frac{1}{k(\xi)} \frac{\partial}{\partial \xi} \left(k(\xi) \frac{\partial w}{\partial \xi} \right) - 2\delta \frac{\partial w}{\partial \tau} + f(\xi, \tau), \quad 0 < \xi < 1, \tau > 0 \\ \left. \frac{\partial w}{\partial \xi} \right|_{\xi=0} &= 0; \quad \left. \left(\frac{\partial w}{\partial \xi} - hv \right) \right|_{\xi=1} = 0 \\ w(\xi, 0) &= 0; \quad \frac{\partial w}{\partial \xi}(\xi, 0) = 0 \end{aligned}$$

Решение задачи а) имеет вид:

$$u(\xi, \tau) = \sum_{n=1}^{\infty} \left(\hat{T}_n \cos \Omega_n \tau + \frac{\tilde{T}_n + \delta \hat{T}_n}{\Omega_n} \sin \Omega_n \tau \right) \varphi_n(\xi) e^{-\delta \tau} \quad (4)$$

где $\{\varphi_n(\xi)\}$ - ортонормированная с весом $k(\xi)$ система собственных функций краевой задачи (1), $\Omega_n = \sqrt{\lambda_n a^2 - \delta^2}$, λ_n - собственные значения краевой задачи, \hat{T}_n, \tilde{T}_n - коэффициенты Фурье начальных условий:

$$\hat{T}_n = (p_0(\xi), \varphi_n) = \int_0^1 p_0(\xi) k(\xi) \varphi_n(\xi) d\xi, \quad \tilde{T}_n = (p_1(\xi), \varphi_n).$$

Решение задачи б) выглядит так:

$$v(\xi, \tau) = a^2 k(0) \sum_{n=1}^{\infty} I_n(\tau) \varphi_n(\xi) \varphi_n(0) e^{-\delta \tau} \quad (5)$$

Здесь

$$I_n(\tau) = I_n^{(1)}(\tau) \left(\cos \Omega_n \tau - \frac{\delta}{\Omega_n} \sin \Omega_n \tau \right) + I_n^{(2)}(\tau) \left(\sin \Omega_n \tau + \frac{\delta}{\Omega_n} \cos \Omega_n \tau \right) - Q(0) \frac{\sin \Omega_n \tau}{\Omega_n}, \quad (6)$$

а

$$I_n^{(1)}(\tau) = \int_0^\tau e^{\delta \eta} \cos \Omega_n \eta d\eta, \quad I_n^{(2)}(\tau) = \int_0^\tau e^{\delta \eta} \sin \Omega_n \eta d\eta. \quad (7)$$

В случае точечного источника фрикативного возбуждения внутри тракта (в точке $\xi = \xi_0$ и с временной амплитудой $\chi(\tau)$), то есть, если $f(\xi, \tau) = \chi(\tau) \delta(\xi - \xi_0)$, решение задачи в) записывается в виде:

$$w(\xi, \tau) = \sum_{n=1}^{\infty} K_n(\tau) \varphi_n(\xi) \varphi_n(\xi_0) e^{-\delta \tau}, \quad (8)$$

где

$$K_n(\tau) = \frac{\sin \Omega_n \tau}{\Omega_n} \int_0^\tau \chi(\eta) e^{\delta \eta} \cos \Omega_n \eta d\eta - \cos \Omega_n \tau \int_0^\tau \chi(\eta) e^{\delta \eta} \sin \Omega_n \eta d\eta. \quad (9)$$

Приведенные формулы для решения справедливы в случае постоянного параметра сопротивления δ . Можно, однако, основываясь на виде полученного решения, сделать допущение о том, что величина δ различна для разных собственных функций (собственных частот) тракта. Математически это означает замену δ на переменную δ_n в формулах (4) – (9) с учетом того, что теперь $\Omega_n = \sqrt{\lambda_n a^2 - \delta_n^2}$.

2.3. Модель турбулентного источника возбуждения

При артикуляции фрикативных звуков в некоторых областях речевого тракта создаются условия для возникновения турбулентности протекающего воздушного потока. В результате этого генерируется шум, имеющий характерный спектр для каждого фрикативного звука. В настоящее время существует лишь так называемая полуэмпирическая теория турбулентных процессов в газовых и жидких средах, т.е. такая теория, в которую входит ряд параметров и зависимостей, устанавливаемых непосредственно опытным путем для конкретного класса систем.

Известно несколько механизмов турбулизации потока в речевом тракте [5]. Первый механизм заключается в том, что турбулентность протекающего воздушного потока возникает, как только число Рейнольдса Re превышает критическую величину Re_{crit} . Число Рейнольдса определяется как $Re(x) = \rho v(x) h(x) / \mu$, где $v(x)$ – линейная скорость потока в точке с координатой x , μ – вязкость воздуха (0.000186 г/смсек), $h(x)$ – характерный геометрический размер сечения трубы. Для трубы постоянного кругового поперечного сечения число $h(x) = const$ соответствует ее диаметру. В экспериментах на механических моделях речевого тракта было найдено, что $Re_{crit} = 1800$. Турбулентный источник, порождаемый этим механизмом, является монополюсным источником. Согласно экспериментальным данным [6], амплитуда спектра этого источника обратно пропорциональна частоте.

Второй механизм возникновения турбулентности заключается в том, что воздушный поток, вытекающий из области, в которой число Рейнольдса превышает критическое, набегаем на препятствие и рассеивается на нем. В результате интерференции падающего и рассеянного шумящих потоков вблизи препятствия возникает новый источник турбулентности, имеющий дипольный характер. В отличие от спектра монополюсного источника, спектр дипольного источника имеет максимумы на частотах $f_n = sh(Re(x_0)) v(x_0) n / h(x_0)$. Здесь $sh(Re)$ – число Струхала, x_0 – координата точечного источника турбулентности, а $n = 1, 2, 3, \dots$. По данным [7], число $sh(Re)$ находится в пределах 0.1 – 0.5 при $10^3 < Re < 10^5$. В наших экспериментах для всех фрикативных считалось, что $sh(Re) = 0.2$ и $n=1,2$. Амплитуду первого обертона источника можно оценить как

$$A_1 = k_1 (Re^2(x_0) - Re_{crit}^2),$$

где k_1 – коэффициент порядка $10^{-5} - 10^{-6}$ в зависимости от расположения источника в речевом тракте. Амплитуда второго обертона примерно в десять раз меньше. Точечный дипольный турбулентный источник для решения задачи в) имел вид:

$$f(\xi, \tau) = \chi(\tau) \delta(\xi - \xi_0), \quad \chi(t) = A_1 \zeta_1(t) + 0.1 A_1 \zeta_2(t).$$

Его временная амплитуда $\chi(t)$ есть комбинация функций $\zeta_1(t), \zeta_2(t)$ – результатов отдельной фильтрации белого шума через два БИХ-фильтра, которые имеют полюсы на частотах $\omega_n = 2\pi f_n$ с коэффициентами затухания $g_n = \pi f_n$ ($n=1,2$). Более сложные источники рассматривались как суперпозиция точечных с различными относительными координатами $\xi_0 = x_0 / l$.

Как видно, модель дипольного турбулентного источника определяется числом Рейнольдса и характерными частотами f_n . Чтобы найти их, необходимо вычислить линейную скорость v потока. Для этого в общем случае следует решить систему нелинейных дифференциальных уравнений, которые описывают аэродинамические свойства речевого тракта [8]. Решение этой громоздкой системы требует больших вычислительных затрат, так что для решения обратных задач в реальном времени требуется иной подход. Он основан на упрощенной модели вычисления линейной скорости [2]. Модель эта, однако, оказалась весьма эффективной как в процессе решения обратной задачи, так и в синтезе фрикативных. Суть ее в следующем.

На низких частотах речевой тракт можно представить эквивалентным электрическим контуром, который состоит из активных сопротивлений R_l , R_{vs} , R_{vt} , соединенных последовательно. Здесь R_l – сопротивление легких и трахеи, R_{vs} – сопротивление голосовой щели, R_{vt} – сопротивление речевого тракта. Тогда поток в тракте w вычисляется как $w = P_l / (R_l + R_{vs} + R_{vt})$, P_l – давление в легких. Каждое сопротивление имеет две компоненты: кинетическую компоненту R_k и компоненту R_v , связанную с потерями на вязкое трение, $R = R_k + R_v$. Кинетическое сопротивление R_k определяется как $R_k = \rho c_k w / 2S^2$, где c_k – коэффициент динамического сопротивления, S – площадь поперечного сечения трубы в месте, где расположено сопротивление. Сопротивление на вязкое трение $R_v = 12\mu h^2 / S^2$. Отсюда получаем уравнение для определения потока w в тракте:

$$0.5w^2 \left(\frac{c_l}{S_l^2} + \frac{c_{vs}}{S_{vs}^2} + \frac{c_{vt}}{S_{vt}^2} \right) + w(R_{vl} + R_{vvs} + R_{vvt}) - P_l = 0. \quad (10)$$

В этом соотношении: c_l – коэффициент динамического сопротивления легких, c_{vs} – коэффициент динамического сопротивления голосовой щели, c_{vt} – коэффициент динамического сопротивления речевого тракта (0.03 – 0.05), R_{vl} – вязкое сопротивление в легких (3 г/сек см⁴), R_{vvs} – вязкое сопротивление в голосовой щели, R_{vvt} – вязкое сопротивление в речевом тракте, S_l – площадь поперечного сечения легких, S_{vs} – площадь голосовой щели, S_{vt} – площадь поперечного сечения речевого тракта в месте наибольшего сужения. Используя весьма грубую модель легких, можно положить член c_l / S_l^2 равным 66 (1/см⁴), $c_{vs} = 0.75 - 0.037S_{vs}$. Площадь сечения голосовой щели S_{vs} вычисляется по формулам, указанным в [1]. Уравнение (10) легко решается относительно w . Линейная скорость потока v определяется по w : $v = w / S_{vt}$.

Рассмотрим также вопрос о виде и количестве источников турбулентности.

Согласно результатам [7], для большинства фрикативных в месте тракта с максимальным числом Рейнольдса присутствуют два источника – монополярный и дипольный. Кроме того, при артикуляции /s, z/ присутствует дополнительный дипольный источник на губах, при артикуляции /sh, zh/ присутствуют два дополнительных дипольных источника – на зубах и губах, а при артикуляции /h/ имеются три дополнительных дипольных источника – на твердом небе, зубах и губах. При артикуляции звуков /f, θ, δ/ участвует только монополярный источник. Кроме того, для всех типов фрикативных имеется дипольный турбулентный источник, расположенный на выходе из голосовой щели. Однако предварительные численные эксперименты показали, что при решении обратной задачи для всех типов фрикативных достаточно использовать в задаче в) один точечный дипольный источник и располагать его в месте верхнего отдела речевого тракта с максимальным числом Рейнольдса. По этой причине дополнительные турбулентные источники в процессе решения обратной задачи не использовались. В экспериментах по синтезу речи выяснилось, что, помимо дипольного источника, расположенного в месте с максимальным числом Рейнольдса, необходимо использовать еще один дополнительный дипольный источник на выходе из голосовой щели. Использование других монополярных и дипольных источников не приводило к улучшению качества синтетических фрикативных.

3. Постановка обратной задачи для фрикативных, алгоритм ее решения и результаты

Речевая обратная задача определяется как задача нахождения функции площади поперечного сечения речевого тракта, или его формы в среднесагиттальном сечении, или артикуляторных параметров, или управлений артикуляторами по измеренным параметрам речевого сигнала. Математически, эта задача решается как задача поиска условного экстремума некоторого критерия оптимальности при различного рода ограничениях. Речевая обратная задача является многоэкстремальной вследствие неоднозначности отображения пространства акустических параметров в артикуляторное пространство. Чаще всего такие задачи решаются многократно с различными начальными приближениями. Это позволяет оценить устойчивость получаемого решения и, если не найти глобальный экстремум, то хотя бы выбрать решение с наиболее подходящими характеристиками.

Выбор начальных приближений сам по себе является непростой задачей. Эта задача может быть решена путем создания так называемой кодовой книги, в которой содержатся векторы артикуляторных параметров и соответствующие им векторы акустических параметров. Использование измерений на установке типа микролучевого рентгеноскопа или электро-магнитных систем (типа Артикулографа) открыло возможность создания кодовых книг, основанных на реальных артикуляторных и акустических данных. Синхронное измерение акустических параметров и координат нескольких точек на языке, губах, верхних и нижних зубах позволяет поставить задачу

формирования кодовой книги для реальных дикторов как специфическую обратную задачу. Решение этой задачи значительно легче, чем решение задачи, когда входными данными служат только акустические параметры, особенно если при этом доступны и измерения формы твердого неба в среднесагиттальной плоскости, формы челюсти в латеральной плоскости, расстояние от передних зубов до задней поверхности тракта и положение гортани.

В данной работе исследовались две обратные задачи для фрикативных звуков. В первой задаче использовались как измеренные акустические параметры спектров фрикативных, так и координаты опорных точек, измеренные с помощью микролучевого рентгеноскопа. Во второй обратной задаче опорные точки не использовались. Решения первой задачи используются для создания кодовой книги начальных приближений. Эта кодовая книга нужна при решении обратной задачи в общем случае, когда координаты опорных точек неизвестны, а имеются только акустические параметры фрикативного. Вторая задача моделирует этот общий случай. И та, и другая задача решалась для речевых данных, в которых фрикативные звуки входили в состав слогов вместе с гласными звуками. При этом для гласных экспериментальной информацией являлись треки формант, а для фрикативных – спектральные данные, описываемые ниже. Сравнение результатов решения обеих обратных задач позволяет оценить устойчивость найденных в них форм речевого тракта.

3.1. Акустические параметры

Спектр фрикативных характеризуется нестабильностью, поэтому, несмотря на относительно успешное решение обратной задачи для фрикативных с использованием спектра энергии, описанное в [2], представляется необходимым поиск устойчивых акустических параметров для фрикативных. Это тем более актуально, что прием речевого сигнала через разные типы микрофонов сопровождается дополнительными существенными искажениями спектров. Влияние типа микрофона иллюстрируется Рис. 1.

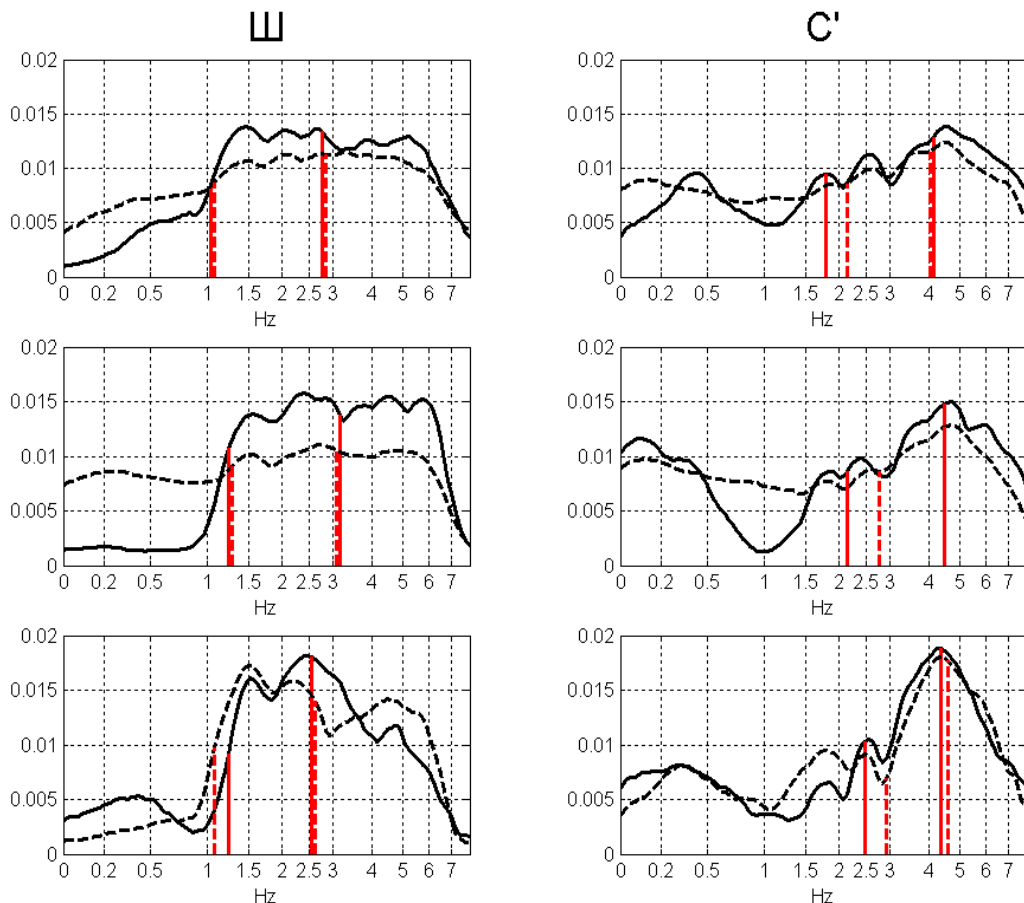


Рис. 1

На нем показаны спектры фрикативных /C/, /Ш/ из слова /шесть/, три раза произнесенного одним и тем же диктором и записанного одновременно парой различных микрофонов. Верхние два рисунка соответствуют усредненным спектрам сигналов, принятых через телефонную трубку (сплошная линия) и направленный микрофон, расположенный на расстоянии примерно в 30 см от рта диктора

(пунктирная линия). Средние два рисунка соответствуют спектрам сигналов, принятых через телефонную трубку другого типа (сплошная линия) и кардиоидный микрофон, укрепленный на мониторе компьютера на расстоянии примерно в 100 см от рта диктора (пунктирная линия). Нижние два рисунка соответствуют спектрам сигналов, принятых через микрофон на головной гарнитуре, расположенный на расстоянии примерно в 5 см от рта диктора (сплошная линия), и кардиоидный микрофон, укрепленный на мониторе компьютера на расстоянии примерно в 100 см от рта диктора (пунктирная линия). Видно, что спектры фрикативных сильно варьируют от произнесения к произнесению и заметно отличаются даже для одного и того же произнесения, но разных типов микрофона. В последнем случае форма речевого тракта одинакова для обоих сигналов, но, в силу склонности решения обратной задачи к неустойчивости, нет гарантии восстановления одной и той же формы тракта. Все это создает трудно преодолимые проблемы при решении обратной задачи и заставляет предпринять поиск более устойчивых параметров спектра фрикативных. Этот поиск выполнялся на обширной базе речевых данных для числительных русского языка, собранной для 47 дикторов, двух типов телефонных трубок и трех типов микрофонов. В этой базе была произведена разметка речевых сигналов на фонетические сегменты, что позволяет выполнять статистический анализ свойств фрикативных звуков.

Так же, как и в [1], энергетический спектр речевого сигнала вычислялся в скользящем окне на 256 точек, и из него вычитался спектр шума, определенный на паузе. Но в данной работе шкала частот в герцах приводилась к шкале частот в мелах и спектр интерполировался кубическими сплайнами. В отличие от задачи для гласных, спектр сглаживался более широкими треугольными равнобедренными фильтрами с полосами пропускания в ± 250 мел. Затем вычислялся логарифм спектра, и частоты в полосе от 0 до 0.2 кГц и от 6 кГц до 8 кГц подавлялись фильтром с линейно падающей характеристикой.

Анализ сглаженных нормированных спектров фрикативных позволяет сделать некоторые предположения о различительных признаках. На Рис. 2 показаны спектры фрикативных, усредненные по большому количеству произнесений и различных условий записи сигнала. Как видно, имеется группа высоких фрикативных типа /С, Ш/, у которых энергия сосредоточена в области высоких частот. В эту же группу входят и мягкие аналоги этих фрикативных, а также фрикативные сегменты аффрикат /Ц, Ч/. Фрикативный /Х/ характеризуется выраженным пиком энергии в области 1300 Гц, а спектр /Ф/ почти плоский. /С/ и /Ш/ различаются шириной спектральной области.

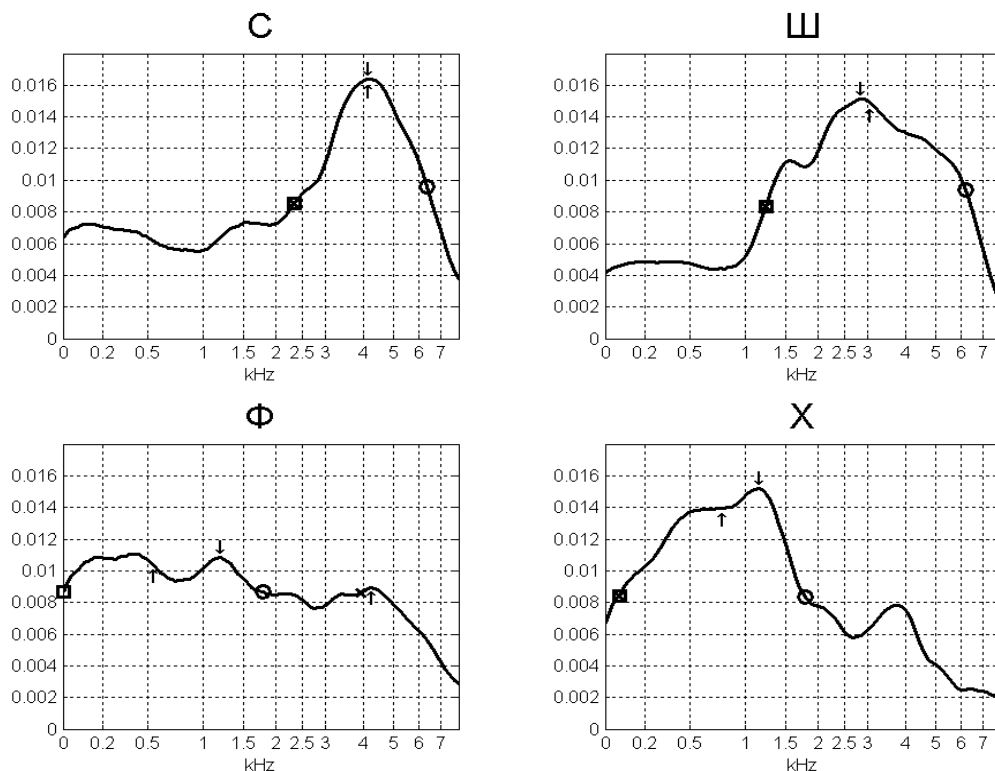


Рис. 2

Средние спектры фрикативных: ↓- глобальный максимум, ↑- центр тяжести, □ - пересечение среднего со стороны низких частот, ○ - пересечение среднего со стороны высоких частот.

Исследовались следующие характерные частоты: частота глобального максимума спектра, частота максимальной положительной производной спектра по частоте, частота первого пересечения среднего уровня и частота центра тяжести при движении по спектру от низких частот к высоким и наоборот. Было установлено, что использование частоты центра тяжести в области энергии выше среднего предпочтительнее, чем частота глобального максимума вследствие меньшего разброса значений. Частота максимальной производной имеет сравнительно большую дисперсию из-за появления резонансов в разных контекстах. Было также найдено, что энергия в области частот ниже 700 Гц сильно зависит от расстояния до микрофона, поэтому во всех случаях эта область спектра не учитывалась.

В результате были отобраны две частоты - частота превышения спектра над средним значением при опросе от высоких частот к низким и частота центра тяжести в этой области. На Рис. 1 эти частоты показаны в виде вертикальных линий. Складывается впечатление, что эти параметры более устойчивы, чем сами спектры. На Рис. 3 показаны распределения этих параметров для фрикативных /С, Ш, Ф, Х/, измеренные на обширной базе данных.

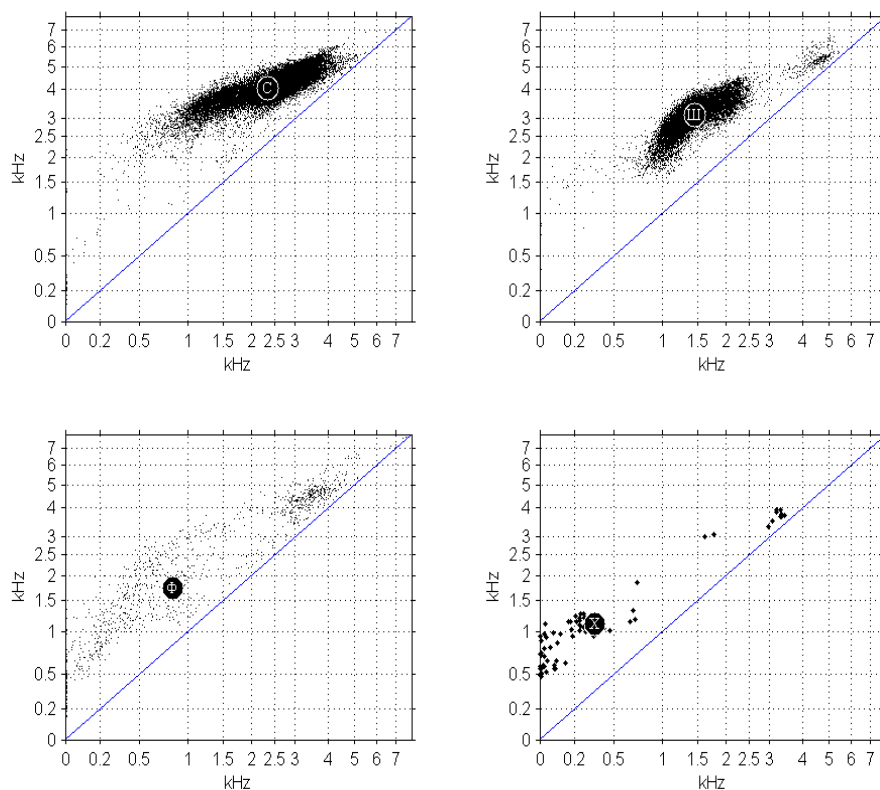


Рис. 3.

Распределение фрикативных. По оси абсцисс – частота пересечения среднего со стороны высоких частот, по оси ординат – частота центра тяжести в области частот с энергией, превышающей среднюю.

Степень пересечения этих двумерных распределений, представленных с точностью 5%, приведена в Табл.1.

Табл. 1.

	С	Ш	Ф	Х
С	100	30	32	5
Ш	30	100	17	4
Ф	32	17	100	24
Х	5	4	24	100

Как видно, фрикативный /Х/ хорошо отделяется по этим параметрам от фрикативных /С, Ш/. Учитывая тот факт, что для /Х/ характерно наличие пика энергии в области 1300 Гц, можно еще улучшить его разделение с высокими фрикативными (Рис. 4).

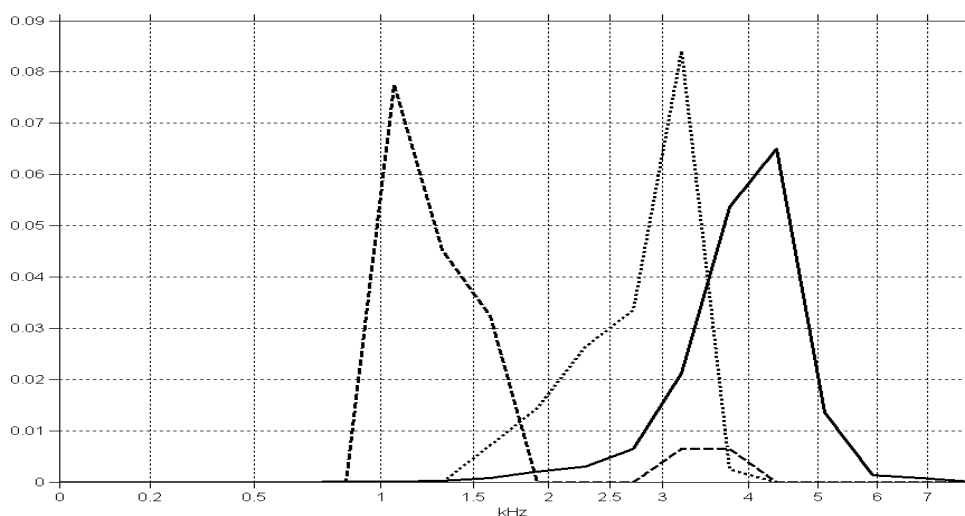


Рис. 4.

Гистограммы распределений частоты глобального максимума спектра для высоких фрикативных /С, С', Ц, Ш, Ш', Ч/ (—), фрикативного /X/ (---) и мягкого фрикативного /X'/ (- - -).

Пересекаемость высоких фрикативных /С, Ш/ с 30% уменьшается до 20%, если вместо частоты пересечения огибающей спектра его среднего использовать частоту пересечения прямой линии, вычисленной как аппроксимация спектра в смысле минимума среднеквадратического отклонения. Высокие фрикативные близки по месту артикуляции, поэтому неудивительно, что для разных условий записи речевого сигнала эти звуки временами оказываются похожими.

Фрикативный /Ф/ не имеет в среднем ярко выраженного максимума спектра или концентрации энергии в какой-либо полосе частот. Этот звук иногда называют диффузным. Количественную меру диффузности можно определить как отклонение спектра $s(f)$ от равномерного:

$$D = \frac{\int_0^{\Theta} |s(f) - s_0| df}{\Theta s_0},$$

где s_0 – среднее значение спектра, Θ – максимальная частота в спектре речевого сигнала. Распределение фрикативных на плоскости «коэффициент наклона линейной аппроксимации - диффузность» показан на Рис. 5.

Наряду с распределениями вышеупомянутых спектральных параметров рассматривались и свойства сильно сглаженных спектров, полученных после вычитания среднего значения. На Рис. 6 показаны средние спектры фрикативных, в которых сохранились только области, превышающие среднее значение нормированного к энергии спектра. После этого вновь сглаживали спектр путем скользящего среднего в окне 600 мел. Как будет показано ниже, в определенных случаях использование таких спектров приводит к удовлетворительному решению обратных задач.

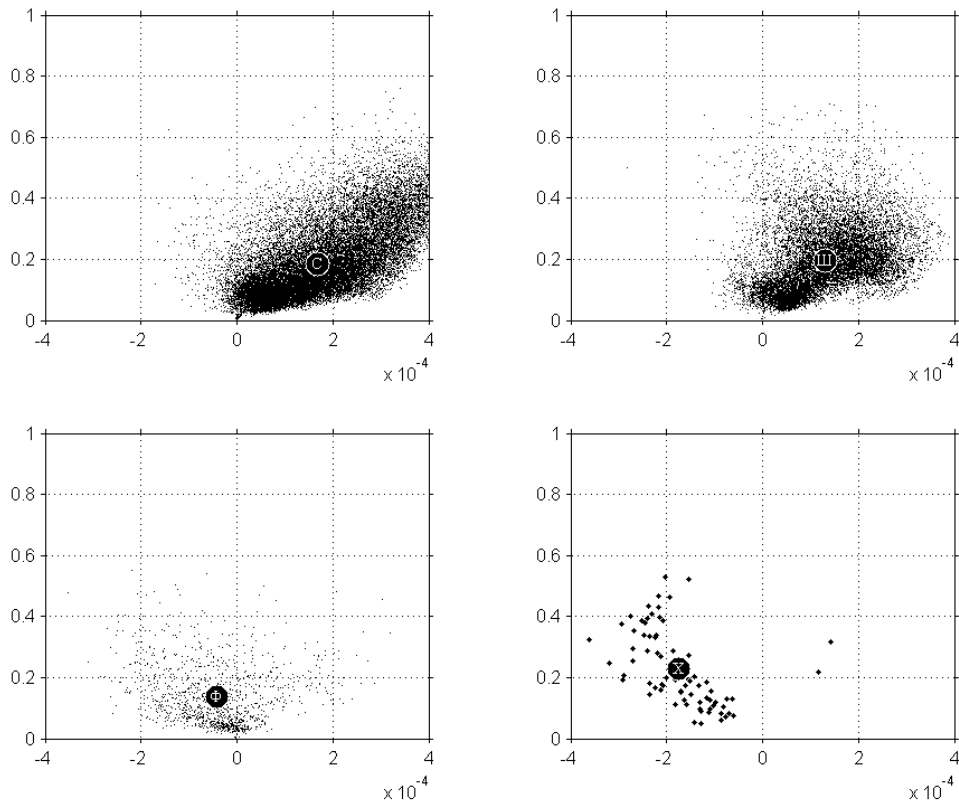


Рис. 5

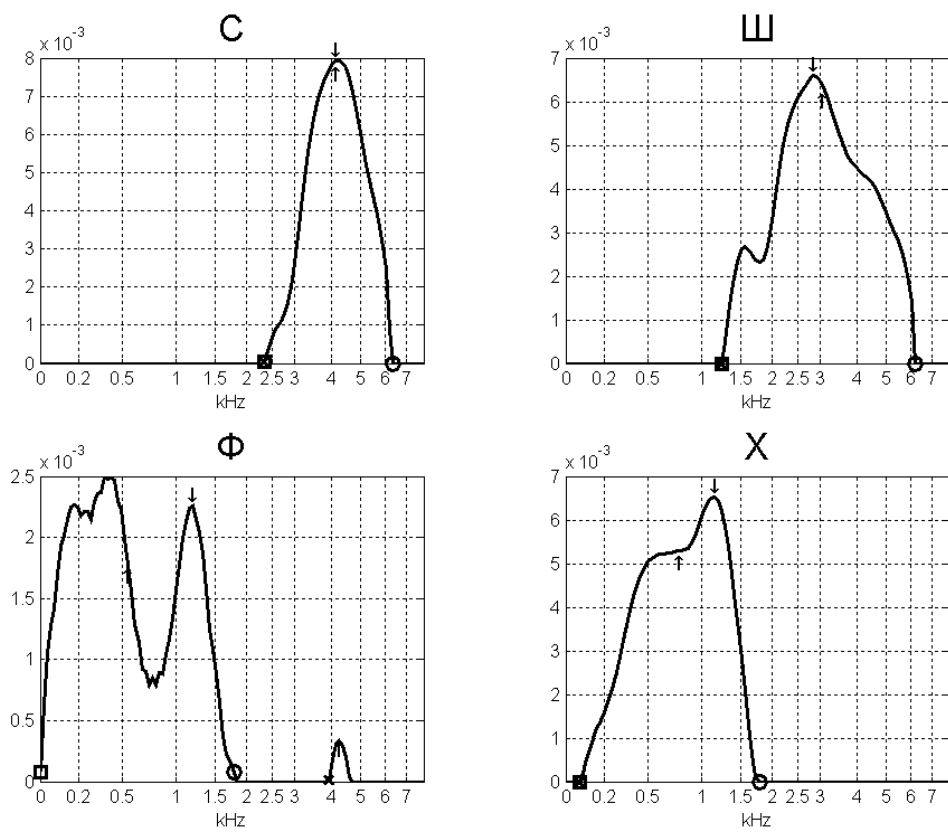


Рис. 6

При аппроксимации распределений в четырехмерном пространстве с координатами “частота пересечения линии аппроксимации, частота центра тяжести спектра в области частот, где спектр выше этой линии, наклон линии и диффузность” несколькими нормальными распределениями степень пересеканости значительно уменьшается (Табл. 2)

Табл. 2.

	С	Ш	Ф	Х
С	100	19	23	3
Ш	19	100	14	3
Ф	23	14	100	24
Х	3	3	24	100

При исследовании обратной задачи для английских фрикативных так же, как и в работе [1], экспериментальной основой послужила база данных, сформированная измерениями на микролучевом рентгенооскопе [9]. В этой базе представлены анатомические параметры 47 дикторов, запись речевого сигнала и синхронные измерения координат 8 точек внутри речевого тракта при произнесении разнообразных текстов. В дальнейшем будем называть эти точки опорными. Решение обратной задачи искалось для фрикативных /s, sh, f, h, z, zh/ в слогах “гласный – фрикативный – гласный” для одного и того же гласного /a/, а для фрикативных /θ, ð/ были вырезаны участки из фразы *the other one* и слова *thing*.

Обратная задача решалась для целого звукосочетания, но на интервалах гласных входными параметрами служили формантные частоты, а на интервале фрикативного – спектральные параметры. Интервал фрикативного определялся вручную.

В соответствии с результатами статистического анализа для русских фрикативных, были вычислены частота пересечения среднего и положение центра тяжести для исследуемых фрикативных английского языка. Эти параметры приведены в Табл. 3.

Табл. 3.

Фрикативный	Частота пересечения среднего Гц	Частота центра тяжести Гц
<i>s</i>	3442	4542
<i>z</i>	3827	4979
<i>sh</i>	1380	2592
<i>zh</i>	1676	2671
<i>f</i>	1474	2536
<i>θ</i>	1474	2494
<i>ð</i>	3535	4391
<i>h</i>	1840	2589

Как видно из этой Таблицы, высокие фрикативные /s, z, sh, zh/ сильно отличаются от диффузных /f, θ, ð/, причем /f, θ/ практически неразличимы на плоскости этих параметров. Это сходство диффузных звуков объективно. Место наибольшего сужения в речевом тракте при образовании звуков /f, θ, ð/ очень близко, и акустические условия генерации турбулентных шумов идентичны – они соответствуют условиям монополюсного источника с падающим спектром и без явно выраженных резонансов.

На Рис. 7 показаны спектры английских фрикативных и линия, аппроксимирующая спектр в полосе от 700 Гц до 6000 Гц. Указаны также коэффициент наклона этой линии и мера диффузности. Видно, что для рассматриваемой реализации губно-зубной фрикативный /f/ отличается от язычно-зубных /θ, ð/ по коэффициенту наклона линейной аппроксимации спектра. Глухой /θ/ и звонкий /ð/ также различаются не только по форме спектра, но и по этим параметрам.

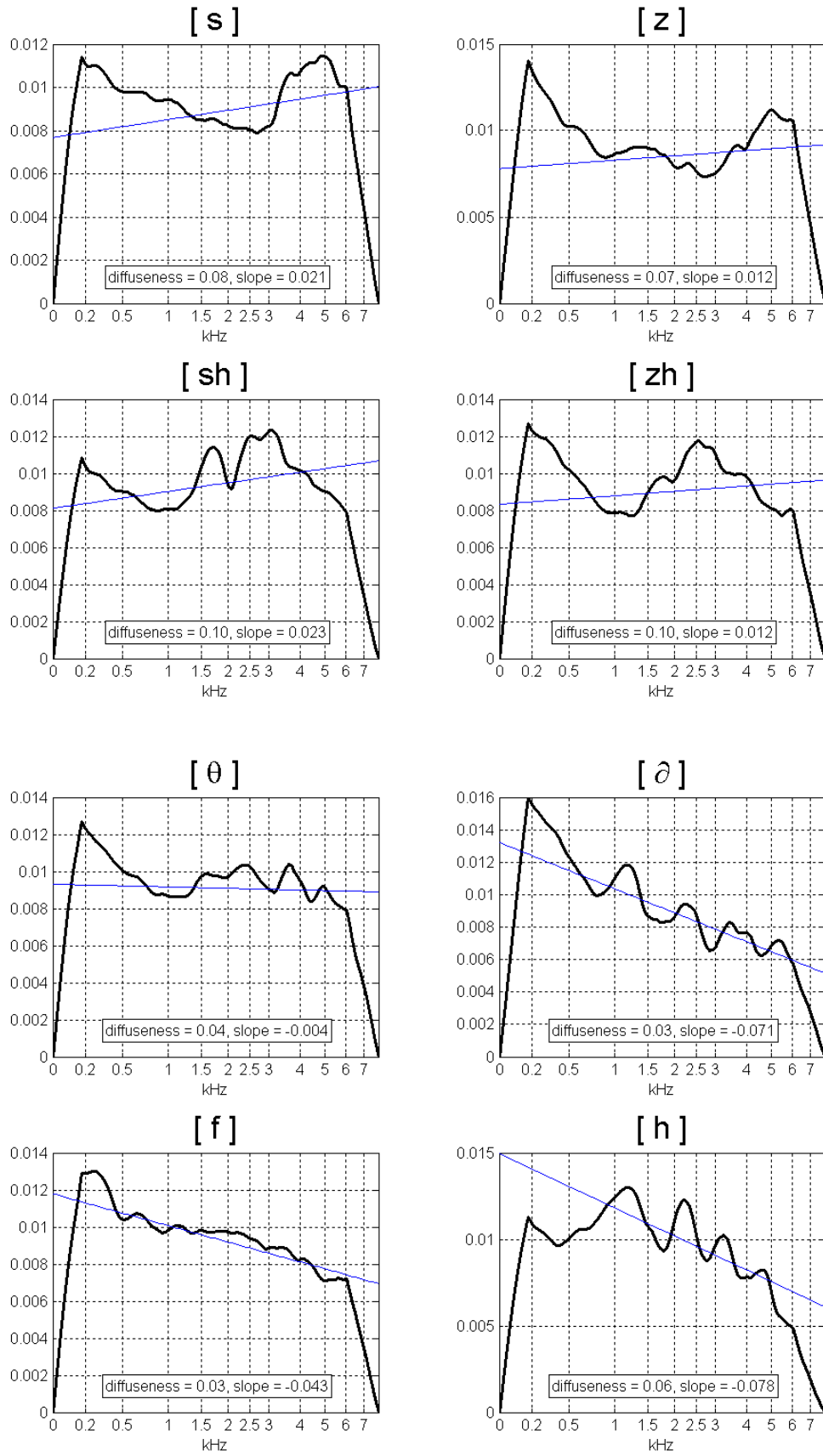


Рис. 7

3.2. Целевые функционалы для решения обратной задачи

Как отмечено выше, обратная задача для фрикативных ставится как задача на условный экстремум. Рассмотрим ее математическую постановку. Обозначим искомые артикуляторные параметры как u^* , целевой функционал, зависящий от вектора этих параметров u - как $\Phi[u]$, а функционалы, описывающие различные ограничения, - как $\Psi_i[u]$. Задача состоит в поиске такого вектора u^* , что

$$\Phi[u^*] = \inf \{ \Phi[u] : \Psi_i[u] \leq 0, i = 1, \dots, N \},$$

(N - число ограничений).

В зависимости от решаемой обратной задачи и типа речевого сегмента могут использоваться различные формы целевого функционала, подлежащего минимизации. Ранее было установлено, что на стационарных участках гласных и фрикативных звуков хорошие результаты получаются при использовании критерия оптимальности в виде работы артикуляторов:

$$\Phi_w(u) = \sum_j c_j (u_j - u_j^0)^2,$$

где j - номер артикуляторного параметра, u_j - артикуляторный параметр, u_j^0 - значение артикуляторного параметра в нейтральном состоянии речевого тракта, c_j - эквивалентная жесткость тканей, присоединенных к соответствующему артикулятору.

На сегментах речи, соответствующих переходным процессам, в дополнение к критерию работы необходимо использовать аналог средней кинетической энергии [10]:

$$\Phi_k(u) = \frac{1}{2T} \sum_j \int_t^{t+T} m_j \left[\frac{du_j(\tau)}{dt} \right]^2 d\tau,$$

где m_j - масса артикулятора, T - интервал времени, на котором оптимизируется процесс управления артикуляцией. В задаче, где заранее не известно, является ли сегмент стационарным или содержит переходный процесс, оба критерия, Φ_w и Φ_k , должны суммироваться с некоторыми весами. Выбор этих весов, по-видимому, зависит от контекста и представляет определенные трудности.

В данной работе исследуется несколько иной вариант композиции критериев оптимальности. Общий целевой функционал задавался в форме: $\Phi[u] = \Phi_1 + \Phi_2 + \alpha \Phi_3$. Первый из составляющих его критериев - Φ_1 применяется, когда использовались данные об опорных внутренних точках речевого тракта. В этом случае

$$\Phi_1 = p_1 \sum_{m=1}^M \text{dist}(P_m^{\text{эксн}}(t), P_m^{\text{выч}}(u(t))) + p_2 \sum_{m=1}^M \left| \frac{d[\text{dist}(P_m^{\text{эксн}}(t), P_m^{\text{выч}}(u(t)))]}{dt} \right|.$$

Здесь $\text{dist}(P_m^{\text{эксн}}(t), P_m^{\text{выч}}(u(t)))$ - евклидово расстояние между экспериментальными опорными точками, измеренными в момент времени t , и соответствующими точками, вычисленными для этого момента; M - число опорных точек. Функционал Φ_1 представляет собой невязку по положению и скорости изменения координат опорных точек. Веса p_1 и p_2 слагаемых этого функционала подбирались экспериментально. Для задачи с неизвестными опорными точками веса считались нулевыми.

Другая составляющая критерия оптимальности - это невязка по акустическим параметрам для гласных, задаваемая как в работе [1]:

$$\Phi_2 = p_3 \max \left\{ \left| \frac{F_k^{\text{выч}}(t)}{F_k^{\text{эксн}}(t)} - 1 \right| : k = 1, 2, 3 \right\}$$

В ней используются измеренные и вычисленные резонансные частоты $F_k(t)$. На фрикативном участке этот функционал заменялся соответствующей мерой сходства экспериментального и вычисленного спектра фрикативного шума. Это были либо расстояния между вычисленными и

измеренными спектрами фрикативного в нормах пространств C и L_2 , либо расстояния между характерными точками этих спектров: частотами пересечения среднего и положения центра масс спектра. В результате численных экспериментов выяснилось, что предпочтительной для сравнения спектров является норма в C , а для сравнения характерных частот лучше использовать векторную норму $\|\cdot\|_\infty$.

В составном критерии оптимизации $\Phi[u]$ всегда использовался дискретный аналог функционала кинетической энергии в виде

$$\Phi_3 = \frac{\|u(t) - u(t - \Delta t)\|_2}{\|u(t - \Delta t)\|_2},$$

где $u(t)$ – вектор артикуляторных параметров в текущий момент времени (на данном «кадре»), $u(t - \Delta t)$ – уже известный вектор параметров на предыдущем кадре, а $\|\cdot\|_2$ – евклидова норма вектора. Смысл этого функционала состоит в минимизации степени изменения артикуляторных параметров. Кроме того, этот функционал служит регуляризующей составляющей в критерии оптимальности $\Phi[u]$, а число α в критерии оптимальности $\Phi[u] = \Phi_1 + \Phi_2 + \alpha \Phi_3$ является параметром регуляризации.

Во всех задачах используются геометрические ограничения $\Psi_i[u]$ на артикуляторные параметры и допустимые формы речевого тракта.

3.3. Метод оптимизации

В силу неоднозначного отображения пространства артикуляторных параметров в пространство акустических параметров, обратные задачи для речевого тракта являются многоэкстремальными. Решение обратной задачи как задачи на поиск экстремума требует использования адекватных алгоритмов оптимизации. Ранее использовались модифицированные методы координатного спуска [2, 11, 12] и сопряженных градиентов [13]. С целью сведения задачи с ограничениями к задаче без ограничений, в этих работах применялась штрафная функция. В настоящее время считается, что для задач нелинейного программирования метод штрафных функций менее эффективен, чем метод, основанный на теореме Куна – Таккера, где ищется седловая точка функции Лагранжа рассматриваемой условной экстремальной задачи. Этот метод успешно применялся для решения обратных задач в [1].

В данной работе использовался алгоритм, основанный на последовательной квадратичной аппроксимации функции Лагранжа и решении соответствующей задачи квадратичного программирования квазиньютоновским методом. Этот алгоритм реализован в стандартном модуле `fminconstr` (или `constr`) пакета MATLAB (Optimization Toolbox). Детали реализации алгоритма изложены в [14]. Русский перевод с комментариями и пояснениями можно найти в [15].

Для решения обратной задачи с помощью этого модуля оптимизации требуется определить начальные приближения. В нашей схеме, как правило, в качестве начального приближения на рассматриваемом временном кадре выбиралось вычисленное решение на предыдущем кадре. На первом кадре начальное приближение выбиралось из кодовой книги, создаваемой в процессе численных экспериментов. В ряде расчетов начальное приближение выбиралось исключительно из кодовой книги.

3.4. Результаты решения обратных задач

Результаты решения двух типов обратных задач для фрикативных показаны в Таблицах 4 и 5. В этих Таблицах приводятся среднеквадратические ошибки реконструкции координат опорных точек и ошибки в равномерной метрике для характерных частот спектра и самого спектра. Напомним, что расстояние в равномерной метрике определяется как максимум модуля разности двух функций. В Таблице 4 показаны результаты решения двух обратных задач, в каждой из которых в качестве акустических данных использовался спектр, но в первой задаче вместе со спектром использовались и координаты опорных точек, а вторая задача решалась только с минимизацией невязки по спектру. В столбце «координаты» приводятся среднеквадратические ошибки восстановления координат опорных точек. В столбце «спектр» показаны расстояния в равномерной метрике между нормированным измеренным и вычисленным спектром, отнесенные к энергии измеренного спектра. В столбце «характерные частоты» приводятся ошибки в равномерной метрике для вычисленных

характерных частот спектра. В столбце «артикуляторные параметры» показана среднеквадратическая ошибка между векторами артикуляторных параметров, вычисленных с использованием разных типов входных данных.

Таблица. 4. Точность решения обратной задачи для акустических данных в виде спектра (%).

фрикативный	спектры и опорные точки			спектры			разница решений артикуляторные параметры
	координаты	спектр	характерные частоты	координаты	спектр	характерные частоты	
<i>s</i>	2.6	30.1	6.8	2.2	34.1	21.2	3.9
<i>sh</i>	2.8	23.8	12.6	2.1	32.6	10.4	3.5
<i>f</i>	3.3	37.9	11.1	2.9	35.3	10.5	1.4
<i>θ</i>	2.8	29.1	8.1	3.0	28.5	7.2	2.3
<i>ð</i>	3.1	30.0	7.6	3.2	29.7	7.7	2.4
<i>h</i>	3.4	37.5	7.9	2.8	35.0	6.8	2.5
<i>z</i>	2.9	39.7	11.5	2.8	32.8	19.6	2.5
<i>zh</i>	2.5	17.3	2.7	3.1	38.1	12.0	9.8

В Таблице 5 показаны результаты решения двух обратных задач, в каждой из которых в качестве акустических данных использовались характерные частоты спектра, но в первой задаче вместе с этими частотами использовались и координаты опорных точек, а вторая задача решалась только с минимизацией невязки по частотам.

Таблица. 5. Точность решения обратной задачи для акустических данных в виде характерных частот спектра (%).

фрикативный	характерные частоты и опорные точки			характерные частоты			разница решений артикуляторные параметры
	координаты	спектр	характерные частоты	координаты	спектр	характерные частоты	
<i>s</i>	1.8	40.0	32.3	2.1	42.8	33.3	3.1
<i>sh</i>	2.9	27.5	13.1	2.3	30.1	11.4	3.6
<i>f</i>	3.8	33.5	13.7	4.3	35.6	21.5	2.7
<i>θ</i>	3.2	30.1	12.8	2.8	29.2	11.5	3.1
<i>ð</i>	3.5	28.7	11.9	2.5	30.1	12.0	2.9
<i>h</i>	4.2	35.4	3.7	3.6	34.1	7.2	2.7
<i>z</i>	2.9	52.6	9.7	2.7	44.7	12.3	2.2
<i>zh</i>	2.8	16.2	2.6	2.9	31.8	8.6	5.7

Из этих Таблиц видно, что решение обратной задачи относительно артикуляторных параметров и координат опорных точек при использовании только акустических параметров достаточно близко к решению обратной задачи с использованием информации и об опорных точках. Точность решения практически не отличается в обоих случаях. Различия между полным набором артикуляторных параметров в обоих случаях также невелики (последний столбец в каждой Таблице), что свидетельствует об устойчивости решения. В то же время, точность восстановления характерных частот спектра оказывается несколько хуже, и еще меньшая точность получена при восстановлении спектров. Это, конечно, связано с нестабильностью спектральных характеристик, присущих фрикативным звукам речи.

Тот факт, что в приведенных результатах не наблюдается разницы в точности решения относительно координат опорных точек при разных способах использования акустических данных – спектров или характерных частот, можно отнести к тому, что исследовались звуко сочетания, записанные только от одного диктора, причем условия записи не изменялись, а отношение сигнал/шум было достаточно хорошим. Следует ожидать, что при разнообразных акустических условиях и для разных дикторов различие в устойчивости описания спектров все-таки проявится, как это наблюдается для обширной базы данных для русского языка.

4. Синтез речи

Синтез звукоочетаний “гласный - фрикатиный - гласный” осуеществлялся путем свертки импульсной характеристики речевого тракта с источником акустических колебаний. На сегментах, соответствующих гласным звукам, в качестве источника голосового возбуждения использовался источник, описанный в [1]. На сегментах, соответствующих глухим фрикатиным, использовался один дипольный источник, расположенный в месте максимального числа Рейнольдса, а также один дипольный источник, расположенный на выходе из голосовой щели. На сегментах, соответствующих звонким фрикатиным, к турбулентным источникам добавлялся голосовой источник.

Результирующий спектр фрикатиных $F(j\omega)$ может быть представлен как

$$F(j\omega) = \sum_{i=1}^N T_i(j\omega)I_i(j\omega)$$

Здесь N – общее количество источников, необходимых для синтеза данного фрикатиного, $I_i(j\omega)$ – спектр i -го источника, приложенного в некоторой координате речевого тракта вдоль средней линии, $T_i(j\omega)$ – передаточная функция речевого тракта, вычисленная в предположении, что тракт возбуждается только i -м источником, а другие источники отсутствуют. Все функции $T_i(j\omega)$ имеют одинаковые полюса. Эти полюса вычисляются с помощью модифицированной схемы длинных линий, учитывающей податливость стенок и разветвления на уровне входа в пищевод [1]. Однако нули функции $T_i(j\omega)$ зависят от расположения i -го источника в речевом тракте. Эти нули совпадают с полюсами текущего импеданса речевого тракта, вычисленного в координате приложенного i -го источника и направленного в сторону голосовой щели. При вычислении нулей предполагалось, что речевой тракт нагружен со стороны голосовой щели на бесконечно большой импеданс. Если i -й источник является голосовым, то в этом случае предполагалось, что $T_i(j\omega)$ содержит только полюса.

На рис. 8 представлены некоторые сонограммы оригинальных и ресинтезированных звукоочетаний "гласный - фрикатиный - гласный". Результаты ресинтеза можно также прослушать. Каждый звуковой файл содержит два речевых сигнала – оригинальный и ресинтезированный. При ресинтезе использовался контур основного тона, заимствованный из оригинального сигнала. Это представляется оправданным, поскольку частота основного тона может быть непосредственно измерена по речевому сигналу и не требует решения обратной задачи.

Для прослушивания необходимо подвести курсор к соответствующему звукоочетанию из нижеследующего списка и нажать левую кнопку мыши:

asa
aza
asha
azha
afa
aha
the other

Для каждой пары слогов на первом месте находится оригинальное звукоочетание, а на втором – ресинтезированное по результатам решения обратной задачи.

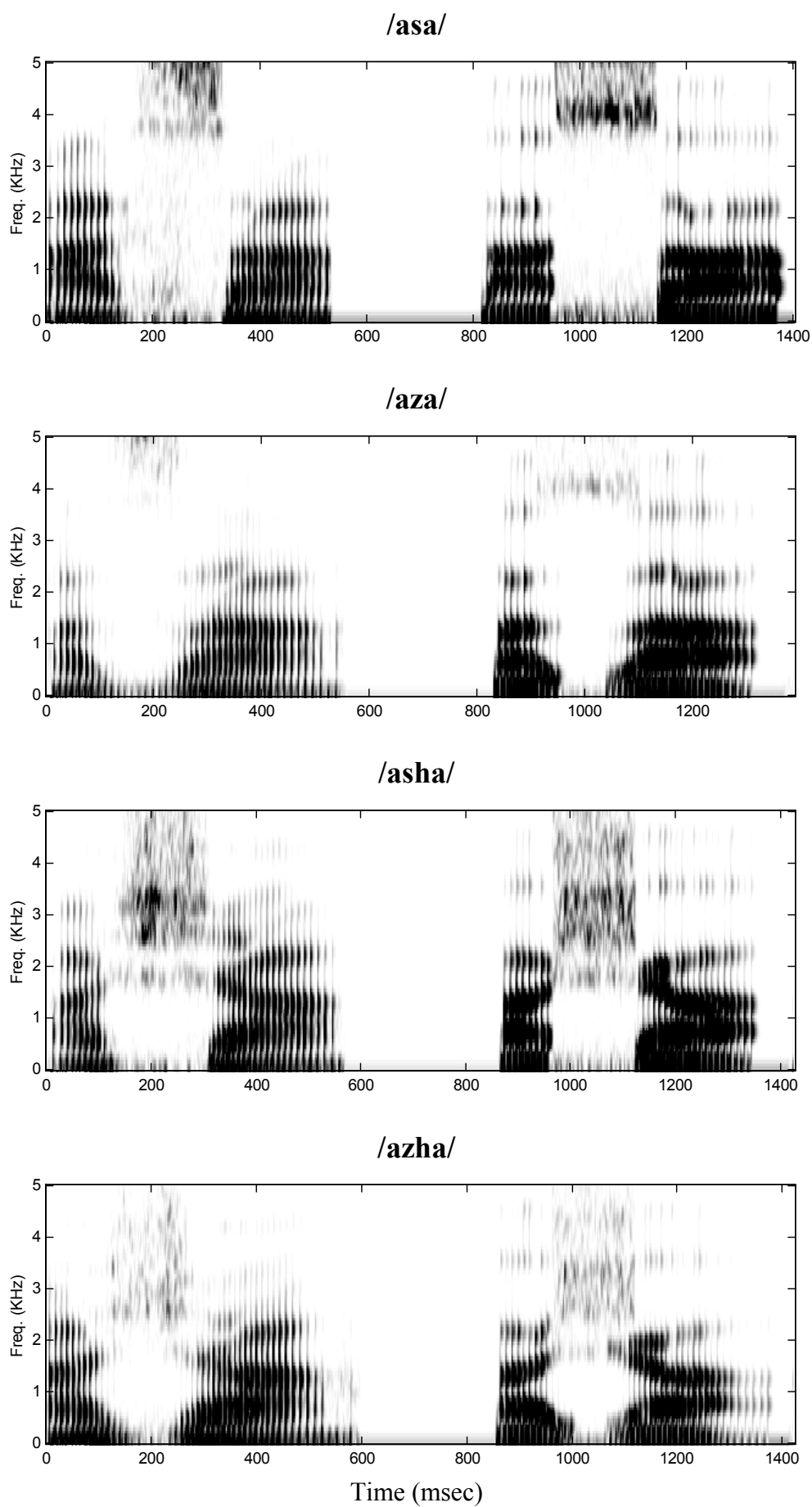


Рис. 8

5. Заключение

Итак, в настоящей работе установлено следующее:

- 1) Решение обратной задачи относительно артикуляторных параметров для фрикативных, когда входными данными служат только акустические параметры речевого сигнала, практически не отличается от решения обратной задачи с информацией в виде акустических данных и треков некоторых точек внутри речевого тракта.
- 2) Решения этой последней задачи можно использовать для формирования кодовой книги начальных приближений, которая необходима в процессе оптимизации при решении общей обратной задачи для речевого тракта.
- 3) Ресинтез фрикативных звуков показывает высокую эффективность используемого алгоритма решения обратных задач и хорошую точность получаемого решения.

Работа выполнена при поддержке РФФИ, грант N03-01-00116.

Список литературы

1. А.С.Леонов, И.С.Макаров, В.Н.Сорокин, А.И.Цыплихин. Артикуляторный ресинтез гласных, Информационные процессы, 2003, т. 3, №2, 73-92.
2. V.N.Sorokin. Inverse problem for fricatives, Speech Communication, 1994, v. 14, N 2, p. 249-262.
3. В.Н.Сорокин. Теория речеобразования. 1985, Радио и связь, М.
4. И.С.Макаров, П.Баден, В.Н.Сорокин. 3-мерная модель речевого тракта и алгоритм вычисления площадей поперечного сечения, Труды Международного семинара "Диалог", 2002, 352-359.
5. K.N.Stevens. Acoustic Phonetics. 2000, MIT Press.
6. L. Pastel. Turbulent voice sources in vocal tract models. 1987, M.S. Thesis, MIT Cambridge.
7. S. Narayanan, A. Alwan. Noise source models for fricative consonants, IEEE Trans. Speech and Audio Processing, 2000, v. 8, №3, pp. 328 - 344.
8. В.Н.Сорокин. Синтез речи. Наука, М., 1992.
9. Westbury J. X-ray Microbeam Speech Production Database. User's Handbook, Version 1, 1994.
10. А.С.Леонов, В.Н.Сорокин. Обратная задача для управления артикуляцией, Доклады Академии Наук, 2000, т. 374, № 6, с. 749-753.
11. V.N.Sorokin. Determination of vocal tract shape for vowels, Speech Communication, 1992, v. 11, N 1, pp. 71-85.
12. V.N.Sorokin, A.V.Trushkin. Articulatory-to-acoustic mapping for inverse problem, Speech Communication, 1996, v. 19, pp. 105-118.
13. V.N.Sorokin, A.S.Leonov and A.V.Trushkin. Estimation of stability and accuracy of inverse problem solution for the vocal tract, Speech Communication, 2000, v. 30, N1, pp. 55-74.
14. Matlab Manual. <http://www.mathworks.com/access/helpdesk/help/toolbox/optim/optim.shtml>.
15. Консультационный центр компании SoftLine. <http://www.matlab.ru/optimiz/index.asp>; http://www.matlab.ru/optimiz/book_1/index.asp.