АКУСТИЧЕСКИЙ ЖУРНАЛ, 2009, том 55, № 2, с. 256–265

АКУСТИКА ЖИВЫХ СИСТЕМ. БИОЛОГИЧЕСКАЯ АКУСТИКА

УДК 612.85

АППРОКСИМАЦИЯ РЕЧЕВОГО ТРАКТА КОНИЧЕСКИМИ РУПОРАМИ

© 2009 г. И. С. Макаров

Институт проблем передачи информации РАН
101447 Москва, Б. Каретный пер. 19
E-mail: speechprod_mak@mail.ru

Поступила в редакцию 18.02.08 г.

Статья посвящена систематическому исследованию метода длинной линии применительно к речевому тракту, аппроксимированному последовательностью конических рупоров. Построенная схема описывает распространение плоских волн в конических рупорах с учетом всех факторов, представляющих интерес для акустической теории речеобразования, – потерь, податливости стенок речевого тракта, а также возможного наличия разветвляющих каналов. Выведенные уравнения протестированы на площадях поперечных сечений речевого тракта, измеренных с помощью магнитно-резонансной томографии на реальном дикторе.

PACS: 43.72.PF, 43.72.FX

ВВЕДЕНИЕ

Для решения многих задач теории речеобразования и речевых технологий необходимо вычислять передаточную функцию речевого тракта и его переносной акустический импеданс по заданной функции его площади поперечного сечения [1, 2]. При этом под передаточной функцией обычно понимается отношение Фурье-образов объемных скоростей на обоих концах тракта (например, на губах и на голосовой щели), а под переносным акустическим импедансом – отношение Фурье-образов давления в некотором сечении речевого тракта и объемной скорости в том же сечении.

Математически задача определения передаточной функции речевого тракта (или его акустического импеданса) сводится к решению спектральной задачи с заданными граничными условиями для уравнения Вебстера:

$$\frac{\partial}{\partial x}\left(S(x)\frac{\partial P}{\partial x}\right) = -S(x)\frac{\omega^2}{c^2}P. \qquad (1)$$

Здесь $S(x)$ – площадь поперечного сечения речевого тракта, x – координата вдоль средней линии тракта, $P(x, j\omega)$ – Фурье-образ звукового давления в тракте, c – скорость звука в речевом тракте, ω – круговая частота (рад/с), $j = \sqrt{-1}$. Это уравнение справедливо в диапазоне частот ниже 4.5 кГц для неизменной во времени площади $S(x)$ [1].

В качестве граничного условия на голосовой щели ($x = 0$) обычно принимается граничное условие второго рода (абсолютно жесткая стенка):

$$\left.\frac{\partial P}{\partial x}\right|_{x=0} = 0. \qquad (2)$$

В качестве граничного условия на губах ($x = L$, L – длина речевого тракта от голосовой щели до губ) принимается граничное условие импедансного типа:

$$P(L, j\omega) = Z_L(j\omega)U(L, j\omega). \qquad (3)$$

Здесь $U(x, j\omega)$ – Фурье-образ объемной скорости в речевом тракте, $Z_L(j\omega)$ – акустический импеданс излучения на губах, обычно определяемый как [3]

$$Z_L(j\omega) = \frac{\rho_0 c}{S_L}\left(\frac{\omega^2 a^2}{2c^2} + j\frac{8\omega a}{3\pi c}\right). \qquad (4)$$

Здесь ρ_0 – плотность воздуха, S_L – площадь излучающего отверстия, a – радиус излучающего отверстия.

В акустической теории речеобразования известно множество алгоритмов решения уравнения (1) [2]. Одним из наиболее популярных методов решения (1) является метод длинной линии. В этом методе речевой тракт аппроксимируется последовательностью N цилиндрических однородных труб. Предполагая, что в тракте отсутствуют разветвления, потери, и принимая стенки тракта абсолютно жесткими, можно получить

следующее матричное уравнение схемы длинной линии [3]:

$$\begin{pmatrix} P_L \\ U_L \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} P_0 \\ U_0 \end{pmatrix}, \tag{5}$$

где

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \prod_{i=1}^{N} \begin{pmatrix} \mathrm{ch}(\gamma l_i) & -\dfrac{\rho_0 c}{S_i}\mathrm{sh}(\gamma l_i) \\ -\dfrac{S_i}{\rho_0 c}\mathrm{sh}(\gamma l_i) & \mathrm{ch}(\gamma l_i) \end{pmatrix}. \tag{6}$$

Здесь l_i – длина i-той цилиндрической секции, S_i – ее постоянная площадь поперечного сечения, $\gamma = j\omega/c$.

Модификации схемы (5), учитывающие разветвления в речевом тракте, различные виды потерь и податливость стенок речевого тракта, построены в [4–7]. Известны также обобщения схемы длинной линии на случай трехмерного речевого тракта [8].

Описанная схема длинной линии имеет, по крайней мере, два недостатка. Во-первых, для достижения надлежащей точности определения передаточной функции и переносного акустического импеданса речевого тракта требуется аппроксимировать его весьма большим количеством цилиндрических секций – от 20 [5] до 44 [9]. Это приводит к большому количеству вычислительных операций.

Во-вторых, аппроксимация речевого тракта цилиндрическими секциями наиболее эффективна для звуков, не имеющих коротких узких щелей в тракте. К таким звукам относятся, прежде всего, гласные звуки. С другой стороны, для фрикативных звуков, акустические характеристики которых сильно зависят от местоположения и размеров сужения [1], аппроксимация тракта цилиндрическими трубами может оказаться неадекватной [10].

Известно, что ступенчатая интерполяция произвольной функции снижает ошибку аппроксимации очень медленно, но повышает сложность вычислений. Поэтому в ряде работ [1, 10] высказано предположение, что использование конических секций вместо цилиндрических для аппроксимации формы речевого тракта может существенно повысить точность либо при заданной точности уменьшить число требуемых секций. В работе [10] использовались два конических рупора для аппроксимации участка максимального сужения в тракте при артикуляции некоторых типов фрикативных согласных. В работе [11] приведены без вывода соотношения длинной линии для волновода, аппроксимированного последовательностью конических рупоров. При этом предполагалось, что разветвления в волноводе отсутствуют, а площади поперечных сечений любой пары рупоров в месте их сочленения одинаковы. Впоследствии эти соотношения были использованы авторами работ [12, 13] для определения передаточных функций речевого тракта без разветвлений.

Настоящая работа посвящена систематическому исследованию метода длинной линии применительно к речевому тракту, аппроксимированному последовательностью конических рупоров. Для этого случая будут выведены матричные уравнения, обобщающие соотношения (5)–(6). На основе этих уравнений будут построены схемы, учитывающие наличие разветвлений в речевом тракте, потерь и податливости стенок тракта.

ТЕОРИЯ

Рассмотрим сначала один конический рупор кругового поперечного сечения, пренебрегая наличием потерь, податливостью его стенок и возможным присутствием разветвляющих рупоров. Пусть x – координата, отложенная вдоль средней линии рупора от одного конца ($x = 0$) до другого ($x = l$), l – длина рупора. Если r – радиус поперечного сечения рупора при $x = 0$, а R – радиус поперечного сечения рупора при $x = l$, то площадь поперечного сечения рупора как функция координаты x определяется как:

$$S(x) = \pi r^2\left(1 + \frac{R-r}{rl}x\right)^2 = S_0(1+\beta x)^2. \tag{7}$$

Здесь $S_0 = \pi r^2$, $\beta = \dfrac{R-r}{rl}$ – коэффициент раствора рупора.

Подставляя (7) в (1), получаем:

$$\frac{d^2P}{dx^2} + \frac{2\beta}{1+\beta x}\frac{dP}{dx} + k^2P = 0. \tag{8}$$

Здесь $k = \omega/c$.

Будем искать решение уравнения (8) в виде [14]:

$$\begin{aligned} P(x, j\omega) &= M(x, j\omega)\exp\left(-\frac{1}{2}\int\frac{2\beta}{1+\beta x}dx\right) = \\ &= M(x, j\omega)\frac{1}{1+\beta x}. \end{aligned} \tag{9}$$

Подставляя (9) в (8), получаем:

$$\frac{d^2M}{dx^2} + k^2M = 0. \tag{10}$$

Решение уравнения (10) записывается как

$$\begin{aligned} M(x, j\omega) &= \rho_0 c(C_1e^{-\gamma x} + C_2e^{\gamma x}), \\ \gamma &= jk. \end{aligned} \tag{11}$$

Здесь C_1 и C_2 – некоторые произвольные константы. Из (9) и (11) получаем решение уравнения (8) в виде

$$P(x, j\omega) = \frac{\rho_0 c}{1+\beta x}(C_1 e^{-\gamma x} + C_2 e^{\gamma x}). \quad (12)$$

Можно показать [15], что Фурье-образы давления $P(x, j\omega)$ и объемной скорости $U(x, j\omega)$ связаны следующим соотношением:

$$S(x)\frac{\partial P}{\partial x} = -j\omega\rho_0 U. \quad (13)$$

Дифференцируя (12) по x и подставляя в (13), получаем:

$$U(x, j\omega) = \frac{cS(x)}{j\omega(1+\beta x)} \times \times\left(C_1 e^{-\gamma x}\left[\gamma + \frac{\beta}{1+\beta x}\right] - C_2 e^{\gamma x}\left[\gamma - \frac{\beta}{1+\beta x}\right]\right). \quad (14)$$

Для определения констант C_1 и C_2 введем следующие граничные условия на обоих концах рупора:

$$P_0 = (0, j\omega), \quad U_0 = U(0, j\omega)$$
$$P_l = P(l, j\omega), \quad U_l = U(l, j\omega). \quad (15)$$

Исключая константы C_1 и C_2 из (12) и (14) и используя соотношения (15) при $x = 0, l$, получаем:

$$\begin{pmatrix} P_l \\ U_l \end{pmatrix} = \begin{pmatrix} \frac{1}{1+\beta l} & 0 \\ \frac{\beta S_0}{\gamma\rho_0 c} & S_0(1+\beta l) \end{pmatrix} \times \times \begin{pmatrix} \mathrm{ch}(\gamma l) & -\rho_0 c\,\mathrm{sh}(\gamma l) \\ -\frac{\mathrm{sh}(\gamma l)}{\rho_0 c} & \mathrm{ch}(\gamma l) \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\frac{\beta}{\gamma\rho_0 c} & \frac{1}{S_0} \end{pmatrix} \begin{pmatrix} P_0 \\ U_0 \end{pmatrix}. \quad (16)$$

Уравнение (16) связывает давление и объемную скорость на одном конце рупора с давлением и объемной скоростью на другом его конце.

Пусть теперь речевой тракт аппроксимируется последовательностью конических рупоров, где l_i – длина i-того рупора, β_i – коэффициент раствора i-того рупора, N – общее количество рупоров. Введем для каждого рупора независимую координату x_i, $0 \le x_i \le l_i$, $S_{0,i}$ – площадь i – го рупора при $x_i = 0$.

Пусть P_i, U_i – Фурье-образы давления и объемной скорости в i-том рупоре, а P_{i+1}, U_{i+1} – Фурье-образы давления и объемной скорости в $(i+1)$-ом рупоре. Тогда для каждого рупора образы давления и объемной скорости на обоих концах связаны соотношением (16). В месте сочленения рупоров выполняются следующие граничные условия:

$$P_i(l_i, j\omega) = P_{i+1}(0, j\omega)$$
$$U_i(l_i, j\omega) = U_{i+1}(0, j\omega). \quad (17)$$

Из этих граничных условий сразу получаем необходимую систему уравнений в векторно-матричной форме:

$$\begin{pmatrix} P_L \\ U_L \end{pmatrix} = \prod_{i=1}^{N} \Phi_i \Gamma_i \mathrm{H}_i \begin{pmatrix} P_g \\ U_g \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} P_g \\ U_g \end{pmatrix}. \quad (18)$$

Здесь P_g, U_g – Фурье-образы давления и объемной скорости на голосовой щели, а P_L, U_L – Фурье-образы давления и объемной скорости на губах. Матрицы Φ_i, Γ_i, H_i определяются следующим образом:

$$\Phi_i = \begin{pmatrix} \frac{1}{1+\beta_i l_i} & 0 \\ \frac{\beta_i S_{0,i}}{\gamma\rho_0 c} & S_{0,i}(1+\beta_i l_i) \end{pmatrix}, \quad (19a)$$

$$\Gamma_i = \begin{pmatrix} \mathrm{ch}(\gamma l_i) & -\rho_0 c\,\mathrm{sh}(\gamma l_i) \\ -\frac{\mathrm{sh}(\gamma l_i)}{\rho_0 c} & \mathrm{ch}(\gamma l_i) \end{pmatrix}, \quad (19b)$$

$$\mathrm{H}_i = \begin{pmatrix} 1 & 0 \\ -\frac{\beta_i}{\gamma\rho_0 c} & \frac{1}{S_{0,i}} \end{pmatrix}, \quad (19c)$$

Уравнение (18) является искомым. Оно обобщает схему длинной линии на случай волновода, аппроксимированного коническими рупорами кругового поперечного сечения. Если положить $\beta = 0$, то уравнение (18) перейдет в уравнение (6).

Определим передаточную функцию речевого тракта как $T(j\omega) = U_L/U_g$, а входной акустический импеданс в речевой тракт со стороны голосовой щели – как $Z_{\text{in}}(j\omega) = P_g/U_g$. Из (18) получаем:

$$T(j\omega) = \frac{1}{A - CZ_L},$$
$$Z_{\text{in}}(j\omega) = \frac{Z_L D - B}{A - CZ_L}, \quad (20)$$

Предположим теперь, что речевой тракт в некоторой точке вдоль средней линии разветвляется одним боковым проходом. Тогда тракт, аппроксимированный последовательностью конических рупоров, может быть разбит на три части – от голосовой щели до точки разветвления (номера рупоров от первого до i-го), от точки разветвле-

ния до губ (номера рупоров от $(i+1)$-го до N-го) и разветвляющий канал, аппроксимированный K рупорами. Тогда связь между Фурье-образами давления и объемной скорости на голосовой щели и в точке $x_i = l_i$ определяется из (18) следующим образом:

$$\begin{pmatrix} P_{lt} \\ U_{lt} \end{pmatrix} = \prod_{s=1}^{i} \Phi_s \Gamma_s \mathrm{H}_s \begin{pmatrix} P_g \\ U_g \end{pmatrix} = \begin{pmatrix} A_{lt} & B_{lt} \\ C_{lt} & D_{lt} \end{pmatrix} \begin{pmatrix} P_g \\ U_g \end{pmatrix}. \quad (21)$$

Аналогично, давление и объемная скорости в точке $x_{i+1} = 0$ и на губах связаны равенствами:

$$\begin{pmatrix} P_L \\ U_L \end{pmatrix} = \prod_{s=i+1}^{N} \Phi_s \Gamma_s \mathrm{H}_s \begin{pmatrix} P_{ut} \\ U_{ut} \end{pmatrix} = \begin{pmatrix} A_{ut} & B_{ut} \\ C_{ut} & D_{ut} \end{pmatrix} \begin{pmatrix} P_{ut} \\ U_{ut} \end{pmatrix}. \quad (22)$$

В соотношениях (21)–(22) матрицы Φ_s, Γ_s, H_s определяются формулами (19) по характеристикам конических рупоров, аппроксимирующих речевой тракт. Символы *lt* и *ut* обозначают акустические характеристики участков речевого тракта от голосовой щели до разветвления и от разветвления до губ, соответственно.

Фурье-образы давления и объемной скорости на обоих концах разветвляющего канала связаны равенствами:

$$\begin{pmatrix} P_{\mathrm{outbif}} \\ U_{\mathrm{outbif}} \end{pmatrix} = \prod_{s=1}^{K} \Phi'_s \Gamma'_s H'_s \begin{pmatrix} P_{\mathrm{inbif}} \\ U_{\mathrm{inbif}} \end{pmatrix} = \begin{pmatrix} A_{\mathrm{bif}} & B_{\mathrm{bif}} \\ C_{\mathrm{bif}} & D_{\mathrm{bif}} \end{pmatrix} \begin{pmatrix} P_{\mathrm{inbif}} \\ U_{\mathrm{inbif}} \end{pmatrix}. \quad (23)$$

В соотношении (23) матрицы Φ'_s, Γ'_s, H'_s определяются формулами (19) по характеристикам конических рупоров, аппроксимирующих разветвляющий канал. Здесь P_{inbif}, U_{inbif} – Фурье-образы давления и объемной скорости в разветвляющем канале в точке разветвления. Например, для назальных и назализованных звуков точка разветвления соответствует началу носового тракта в районе мягкого неба. P_{outbif}, U_{outbif} – Фурье-образы давления и объемной скорости на другом конце разветвляющего канала. Например, для назальных и назализованных звуков P_{outbif}, U_{outbif} совпадают с образами давления и объемной скорости на излучающей стороне назального тракта (у ноздрей). Символ bif обозначает акустические характеристики речевого тракта в разветвляющем канале.

В точке разветвления имеем следующие граничные условия:

$$\begin{aligned} P_{lt} &= P_{ut} = P_{\mathrm{inbif}} \\ U_{lt} &= U_{ut} + U_{\mathrm{inbif}}. \end{aligned} \quad (24)$$

Из определения акустического импеданса и первой цепочки равенств (24) следует, что $U_{ut} = P_{ut}/Z_{ut} = P_{lt}/Z_{ut}$, $U_{\mathrm{inbif}} = P_{\mathrm{inbif}}/Z_{\mathrm{inbif}} = P_{lt}/Z_{\mathrm{inbif}}$.

Тогда $U_{ut} = U_{lt} - P_{lt}/Z_{\mathrm{inbif}}$, $U_{\mathrm{inbif}} = U_{lt} - P_{lt}/Z_{ut}$. Эти равенства дают возможность объединить уравнения (21) и (22):

$$\begin{pmatrix} P_L \\ U_L \end{pmatrix} = \begin{pmatrix} A_{ut} & B_{ut} \\ C_{ut} & D_{ut} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\dfrac{1}{Z_{\mathrm{inbif}}} & 1 \end{pmatrix} \times \begin{pmatrix} A_{lt} & B_{lt} \\ C_{lt} & D_{lt} \end{pmatrix} \begin{pmatrix} P_g \\ U_g \end{pmatrix}. \quad (25a)$$

Аналогично объединяем уравнения (21) и (23):

$$\begin{pmatrix} P_{\mathrm{outbif}} \\ U_{\mathrm{outbif}} \end{pmatrix} = \begin{pmatrix} A_{\mathrm{bif}} & B_{\mathrm{bif}} \\ C_{\mathrm{bif}} & D_{\mathrm{bif}} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\dfrac{1}{Z_{ut}} & 1 \end{pmatrix} \times \begin{pmatrix} A_{lt} & B_{lt} \\ C_{lt} & D_{lt} \end{pmatrix} \begin{pmatrix} P_g \\ U_g \end{pmatrix}. \quad (25b)$$

Входной акустический импеданс Z_{inbif} в разветвляющий канал со стороны точки разветвления определяется, согласно (20), как

$$Z_{\mathrm{inbif}}(j\omega) = \frac{Z_{\mathrm{rad}} D_{\mathrm{bif}} - B_{\mathrm{bif}}}{A_{\mathrm{bif}} - C_{\mathrm{bif}} Z_{\mathrm{rad}}}. \quad (26)$$

Здесь Z_{rad} – импеданс нагрузки разветвляющего канала.

Аналогично, входной акустический импеданс Z_{ut} в речевой тракт со стороны точки разветвления в направлении губ определяется соотношением вида (20):

$$Z_{ut}(j\omega) = \frac{Z_L D_{ut} - B_{ut}}{A_{ut} - C_{ut} Z_L}. \quad (27)$$

Формулы (25a)–(25b) дают возможность определить передаточную функцию речевого тракта с

разветвлениями. Производя перемножение матриц, получаем:

$$T(j\omega) = \frac{U_L + U_{\text{outbif}}}{U_g} =$$
$$= \frac{1}{(Z_{ut} + Z_{\text{inbif}})A_{lt} - Z_{ut}Z_{\text{inbif}}C_{lt}} \times \tag{28}$$
$$\times \left[\frac{Z_{\text{inbif}}}{A_{ut} - C_{ut}Z_L} + \frac{Z_{ut}}{A_{\text{bif}} - C_{\text{bif}}Z_{\text{rad}}}\right].$$

Рассмотрим теперь податливость стенок, а также наличие потерь на вязкое трение и теплопроводность в речевом тракте, аппроксимированном последовательностью конических рупоров. Возможно несколько способов учета этих факторов в уравнениях (18)–(19). Мы воспользуемся моделью потерь и податливости стенок, построенной в работах [15, 16]. Согласно данной модели, уравнения, связывающие Фурье-образ давления и объемной скорости в речевом тракте с потерями и податливыми стенками, могут быть записаны следующим образом:

$$\frac{\partial U}{\partial x} + \frac{S(x)}{\rho_0 c^2}[\chi(j\omega) + \alpha(j\omega) + j\omega]P = 0,$$
$$\frac{\partial P}{\partial x} + \frac{\rho_0}{S(x)}[\alpha(j\omega) + j\omega]U = 0. \tag{29}$$

Здесь $\alpha(j\omega) = \sqrt{C_{vt}j\omega}$ – функция, учитывающая потери на трение и теплопроводность, C_{vt} – константа, подбираемая эмпирическим путем. Функция $\chi(j\omega)$ вводится для учета податливости стенок речевого тракта. Предполагая, что радиальное смещение стенок речевого тракта локально описывается уравнением гармонического осциллятора, получаем следующее выражение для $\chi(j\omega)$ [15]:

$$\chi(j\omega) = \frac{j\omega\omega_0^2}{\left(j\omega + \frac{R_w}{L_w}\right)j\omega + \frac{1}{C_w L_w}}. \tag{30}$$

Здесь ω_0 – низшая частота радиального резонанса речевого тракта, R_w, L_w, C_w – потери на стенках, масса и податливость стенок тракта на единицу площади, соответственно.

Выражая из уравнений (29) давление $P(x, j\omega)$, получаем следующий эквивалент уравнения Вебстера (1):

$$\frac{\partial}{\partial x}\left(S(x)\frac{\partial P}{\partial x}\right) - \frac{\sigma^2}{c^2}S(x)P = 0,$$
$$\sigma = \sqrt{(\alpha + j\omega)(\chi + \alpha + j\omega)}. \tag{31}$$

Подставляя в это уравнение площадь конического рупора (7) и дословно повторяя все выкладки, предшествовавшие уравнениям (19), получаем следующий аналог этих соотношений для речевого тракта с потерями и податливыми стенками:

$$\Phi_i = \begin{pmatrix} \frac{1}{1 + \beta_i l_i} & 0 \\ \frac{\beta_i S_{0,i}}{\rho_0(\alpha + j\omega)} & S_{0,i}(1 + \beta_i l_i) \end{pmatrix}, \tag{32a}$$

$$\Gamma_i = \begin{pmatrix} \text{ch}\left(\frac{\sigma l_i}{c}\right) & -\rho_0 c\zeta\,\text{sh}\left(\frac{\sigma l_i}{c}\right) \\ -\frac{1}{\rho_0 c\zeta}\text{sh}\left(\frac{\sigma l_i}{c}\right) & \text{ch}\left(\frac{\sigma l_i}{c}\right) \end{pmatrix}, \tag{32b}$$

$$\text{H}_i = \begin{pmatrix} 1 & 0 \\ -\frac{\beta_i}{\rho_0(\alpha + j\omega)} & \frac{1}{S_{0,i}} \end{pmatrix}, \tag{32c}$$

где

$$\zeta = \sqrt{\frac{\alpha(j\omega) + j\omega}{\chi(j\omega) + \alpha(j\omega) + j\omega}}. \tag{33}$$

Для абсолютно жестких стенок при отсутствии потерь $\alpha = \chi = 0$, так что уравнения (32) переходят в уравнения (19). При $\beta_i = 0$ (тракт аппроксимируется последовательностью цилиндрических труб) уравнения (32) переходят в уравнения, выведенные в [5].

Аналогичным образом потери и податливость стенок речевого тракта могут быть введены в разветвляющие каналы.

ЭКСПЕРИМЕНТЫ

Тестирование построенной схемы длинной линии проводилось путем вычисления акустических характеристик речевого тракта по его конической и цилиндрической аппроксимациям с их последующим сравнением. В качестве акустических характеристик использовались резонансные и антирезонансные частоты тракта в диапазоне до 3100 Гц, а также огибающие самих передаточных функций. При этом резонансные (антирезонансные) частоты определялись по максимумам (минимумам) огибающих соответствующих передаточных функций и сравнивались в евклидовой метрике. Для количественной оценки различия между огибающими передаточных функций ис-

пользовалась мера спектрального искажения $\varepsilon(T_{con}(j\omega), T_{cyl}(j\omega))$ в дБ [17]:

$$\varepsilon(T_{con}(j\omega), T_{cyl}(j\omega)) = \\ = \sqrt{\frac{1}{\omega_2 - \omega_1}\int_{\omega_1}^{\omega_2}\left[10\log_{10}\left(\frac{|T_{con}(j\omega)|^2}{|T_{cyl}(j\omega)|^2}\right)\right]^2 d\omega}. \quad (34)$$

Здесь $T_{con}(j\omega)$ – передаточная функция, вычисленная по конической аппроксимации формы речевого тракта; $T_{cyl}(j\omega)$ – передаточная функция, вычисленная по аппроксимации такой же формы речевого тракта последовательностью цилиндрических секций; $\omega_1...\omega_2$ – полоса частот, в которой вычисляются передаточные функции.

В качестве тестового материала использовались площади поперечных сечений, измеренные с помощью магнитно-резонансной томографии, которые были заимствованы из работы [18]. Исходный набор данных представлял функции площади при артикуляции 4-х английских гласных /a, æ, i, u/ в произнесении двух дикторов – всего восемь функций площадей. Каждая функция была представлена как последовательность из 20 отсчетов.

На базе этих площадей был подготовлен новый набор, включавший в себя 16 функций площадей поперечных сечений. При этом 8 функций нового набора совпадали с исходным набором площадей, а 8 оставшихся функций площади были получены путем шунтирования каждой исходной площади назальным трактом. Функция площади назального тракта заимствовалась из работы [19]. Площадь входа в назальный тракт со стороны речевого тракта полагалась равной 0.05 см^2. Назальный тракт шунтировал речевой тракт на расстоянии 9 см от голосовой щели.

При вычислении передаточных функций $T_{con}(j\omega)$ и $T_{cyl}(j\omega)$ учитывалось наличие потерь на вязкое трение и теплопроводность и податливость стенок речевого тракта. Использовались следующие значения констант, входящих в соотношения (29)–(30): $C_{vt} = 4$ (рад/с), $\omega_0 = 600\pi$ (рад/с), $R_w/L_w = 130\pi$ (рад/с), $1/(C_wL_w) = (30\pi)^2$ (рад/с)^2. Для назального тракта низший радиальный резонанс полагался равным $2\omega_0$.

При вычислении $T_{cyl}(j\omega)$ речевой тракт аппроксимировался последовательностью из 19 цилиндрических труб одинаковой длины. Для аппроксимации речевого тракта коническими рупорами исходная функция площади поперечного сечения сначала пересчитывалась в функцию эквивалентного радиуса по формуле $r(x) = \sqrt{S(x)/\pi}$. Поскольку в коническом рупоре радиус является линейной функцией координаты x, функция $r(x)$ аппроксимировалась последовательностью 19 линейных кусков одинаковой длины. Предполагалось, что в месте сочленения двух линейных кусков функция $r(x)$ непрерывна. При этом длина каждого линейного куска полагалась равной длине трубы при цилиндрической аппроксимации. Затем каждый линейный кусок пересчитывался обратно в площадь. В результате исходная функция площади оказывалась аппроксимированной последовательностью 19 функций вида (7). Среднеквадратическая ошибка такой аппроксимации составила доли процента. Аналогичный подход был использован для аппроксимации назального тракта последовательностью конических секций.

По различным данным [20, 21], дифференциальный порог восприятия формантных частот составляет до 5% от частоты первых трех формант. Для неразветвленных площадей невязки между первыми тремя формантными частотами, вычисленными по цилиндрической и конической аппроксимации, практически для всех гласных оказались по абсолютному значению менее 4%, т.е. в пределах дифференциального порога. Исключение составил гласный /u/ в произнесении первого диктора, у которого невязка по частоте второй форманты превысила 13%. Причины такой большой ошибки будут исследованы ниже.

На рис. 1 в качестве примера представлены огибающие передаточных функций $T_{con}(j\omega)$ и $T_{cyl}(j\omega)$ для первого диктора. На каждом рисунке указано значение спектрального искажения (34). Среднее значение величины спектрального искажения для обоих дикторов составило 2.5 дБ. Напомним, что разница между двумя спектрами менее одного децибелла не воспринимается человеком на слух. Максимальное спектральное искажение было равно 5.9 дБ для первого диктора, гласный /u/.

Была проверена гипотеза, согласно которой различие между $T_{con}(j\omega)$ и $T_{cyl}(j\omega)$ может быть связано с разными условиями сочленений секций в цилиндрической и конической аппроксимациях. В первом случае соседние цилиндрические трубы имеют, вообще говоря, разные площади поперечного сечения, так что в месте их сочленения наблюдается скачок площади. Во втором случае площадь тракта меняется непрерывно при переходе от одного рупора к другому. Для проверки этой гипотезы коэффициенты β_i в (32) умножались на коэффициент h, который варьировался в диапазоне от единицы до нуля. Чем меньше h, тем выше амплитуда скачка в месте сочленения конических рупоров. При $h = 0$ конусы вырождаются в цилиндрические секции.

Выяснилось, что для всех функций площадей речевого тракта $T_{con}(j\omega)$ равномерно стремится к $T_{cyl}(j\omega)$ при $h \longrightarrow 0$. На рис. 2 в качестве примера представлены огибающие $T_{con}(j\omega)$ (гласный /u/, первый диктор, см. рис. 1) для двух значений па-

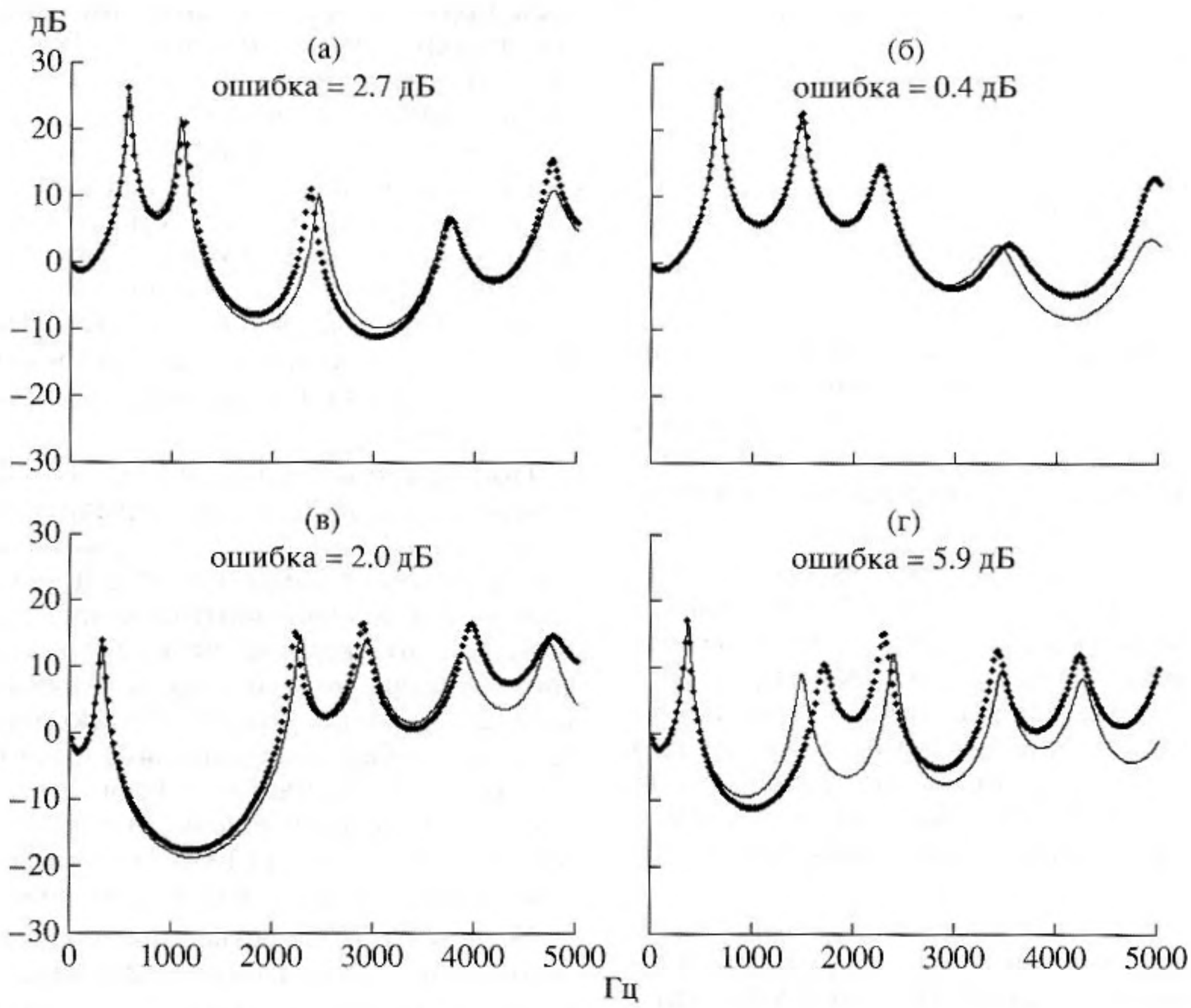

Рис. 1. Огибающие передаточных функций для первого диктора, неразветвленный случай (сплошная линия – коническая аппроксимация, пунктир – цилиндрическая аппроксимация).

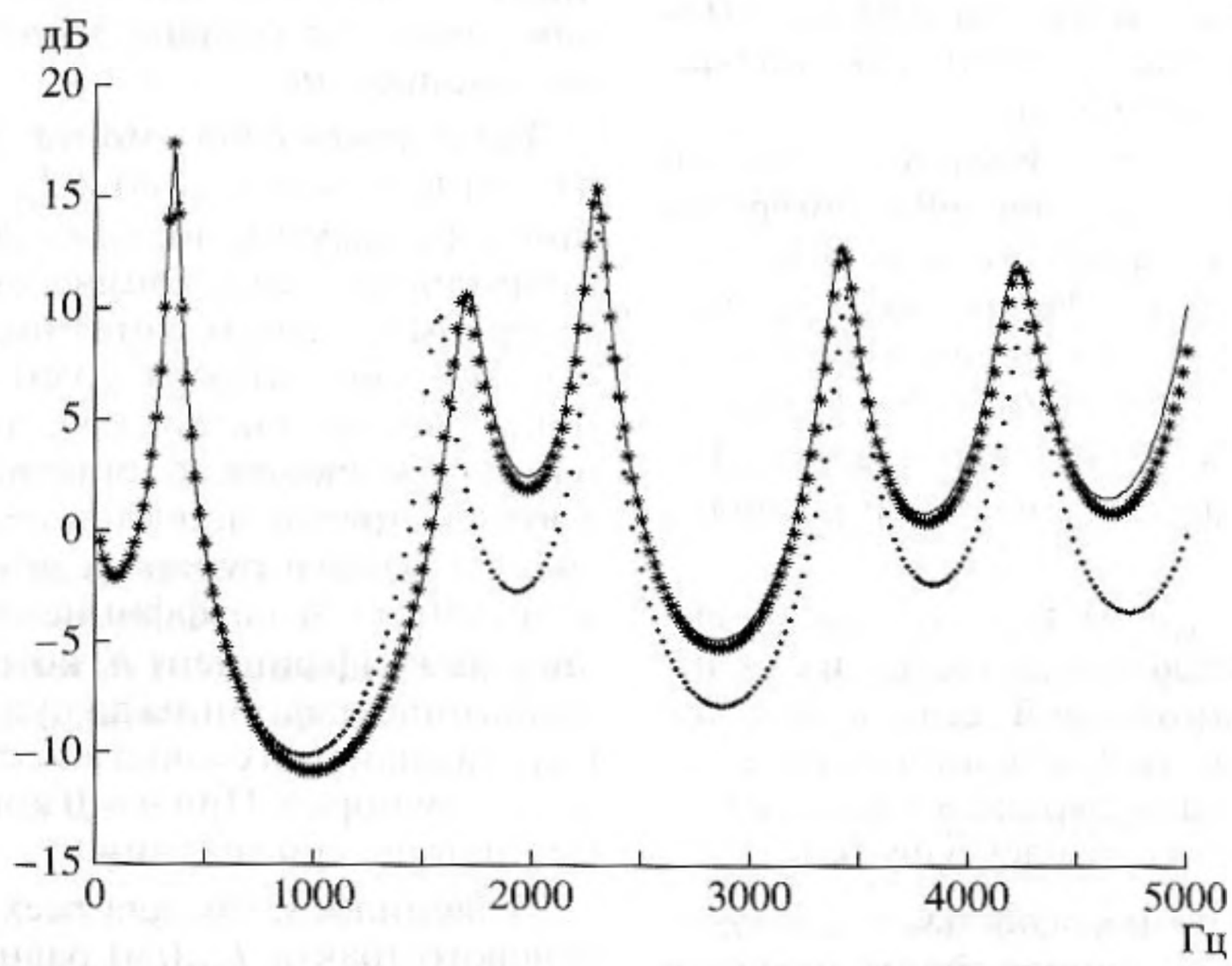

Рис. 2. Огибающие передаточных функций для разных условий сочленения конических секций (сплошная линия – цилиндрическая аппроксимация, пунктир – коническая аппроксимация при $h = 0.9$, пунктир со звездочками – коническая аппроксимация при $h = 0.1$).

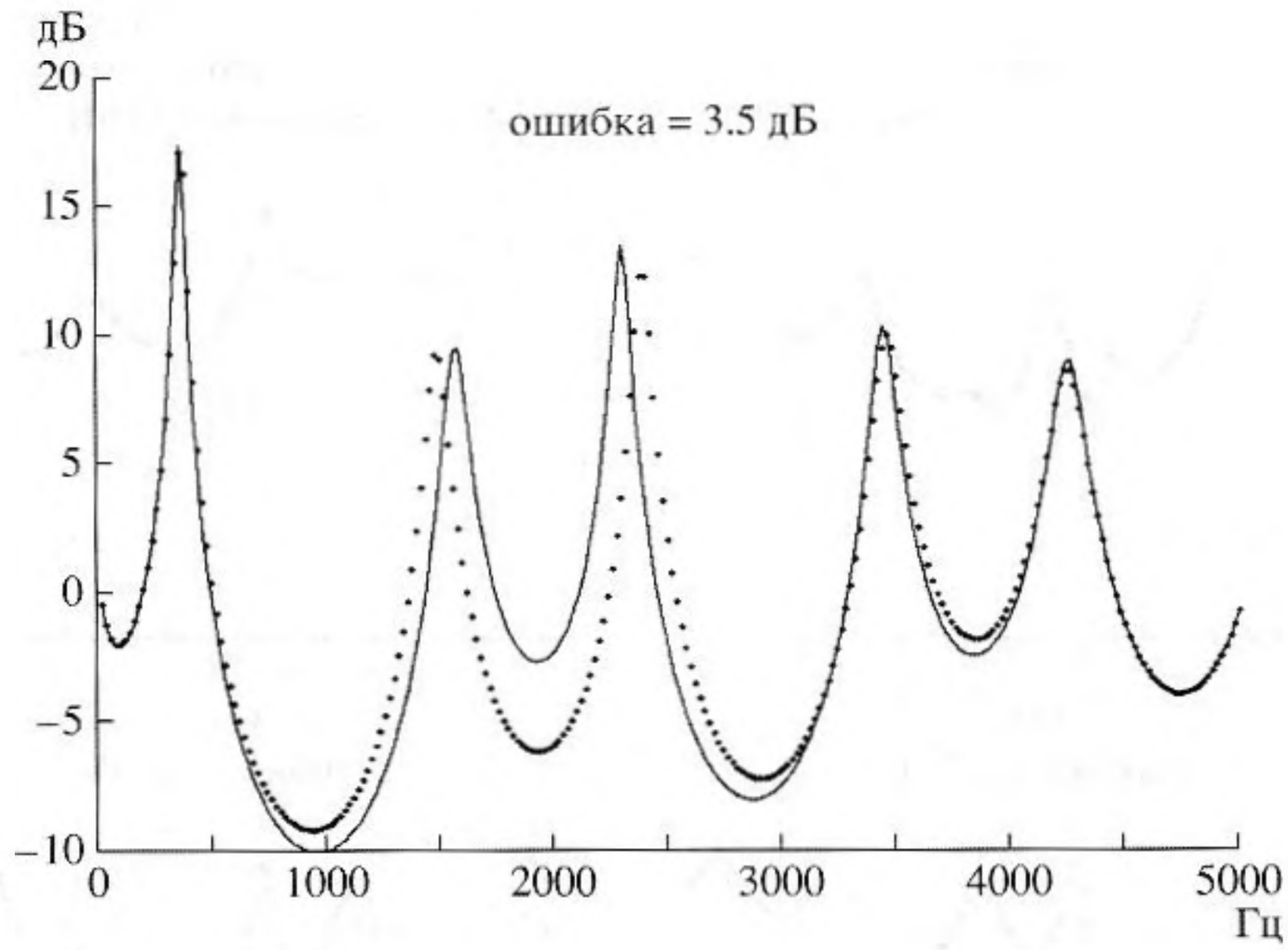

Рис. 3. Огибающие для конической аппроксимации (19 секций, пунктир) и цилиндрической аппроксимации (200 секций, сплошная линия).

раметра h: $h = 0.9$ и $h = 0.1$. Видно, что при $h = 0.1$ $T_{con}(j\omega)$ и $T_{cyl}(j\omega)$ практически идентичны. Ошибка по второй формантной частоте при $h = 0.1$ составляет доли процента.

Таким образом, различие между огибающими $T_{con}(j\omega)$ и $T_{cyl}(j\omega)$ и соответствующими резонансными частотами связано с наличием скачков большой амплитуды в площади поперечного сечения тракта при цилиндрической аппроксимации. Возможны, по крайней мере, два способа компенсации этих скачков. Первый способ заключается в аппроксимации исходной функции площади большим количеством цилиндрических секций. Оказалось, что для всех конфигураций $T_{cyl}(j\omega)$ равномерно стремилась к $T_{con}(j\omega)$ при увеличении количества труб. Вместе с тем, уменьшение спектрального искажения (34) происходило очень медленно.

На рис. 3 в качестве примера показана огибающая $T_{cyl}(j\omega)$, вычисленная по речевому тракту, который аппроксимирован 200-ми цилиндрическими секциями (гласный /u/, первый диктор, см. рис. 1). Из сравнения рис. 1 и рис. 3 видно, что спектральное искажение уменьшается на 2.4 дБ. Соответствующая среднеквадратическая ошибка по второй формантной частоте составляет около 5% (по сравнению с 13% для исходной аппроксимации 19-ю цилиндрическими секциями).

Второй способ заключается в увеличении длины соседних цилиндрических секций с сильно различными площадями на некоторый процент. Различные модели коррекции длин секций построены в [22, 23]. Наши эксперименты показали, что величина этой поправки сильно зависит от конфигурации речевого тракта. В целом, нам не удалось подобрать корректирующий коэффициент длины таким образом, чтобы для каждой функции площади величина спектрального искажения уменьшилась более, чем на 1 дБ. Возможно, что, помимо коррекции длины цилиндрических секций с сильно различающимися площадями, необходимо вносить поправки, зависящие от частоты, в соответствующие сомножители матричного уравнения длинной линии [24]. Этот вопрос требует отдельного исследования.

Рассмотрим теперь результаты вычислений для функций площади поперечных сечений, разветвленных назальным трактом. Невязка между антирезонансными частотами, вычисленными по обеим схемам длинной линии, оказалась весьма малой для всех исследованных конфигураций. По абсолютному значению она не превышала 3%. Невязки по резонансным частотам при наличии назального разветвления также оказались небольшими. По абсолютному значению они не превышали 2.8% практически для всех гласных. Исключение составила конфигурация тракта для гласного /u/ в произнесении первого диктора (невязка по 3-й резонансной частоте >8%).

На рис. 4 в качестве примера представлены огибающие передаточных функций $T_{cyl}(j\omega)$ и $T_{con}(j\omega)$ для речевого тракта первого диктора, разветвленного назальным каналом. Средняя величина спектрального искажения для обоих дикторов составила 2.2 дБ, что сопоставимо со средним спектральным искажением для неразветвленного тракта. Максимальное значение спектрального искажения по-прежнему наблюдалось на гласном

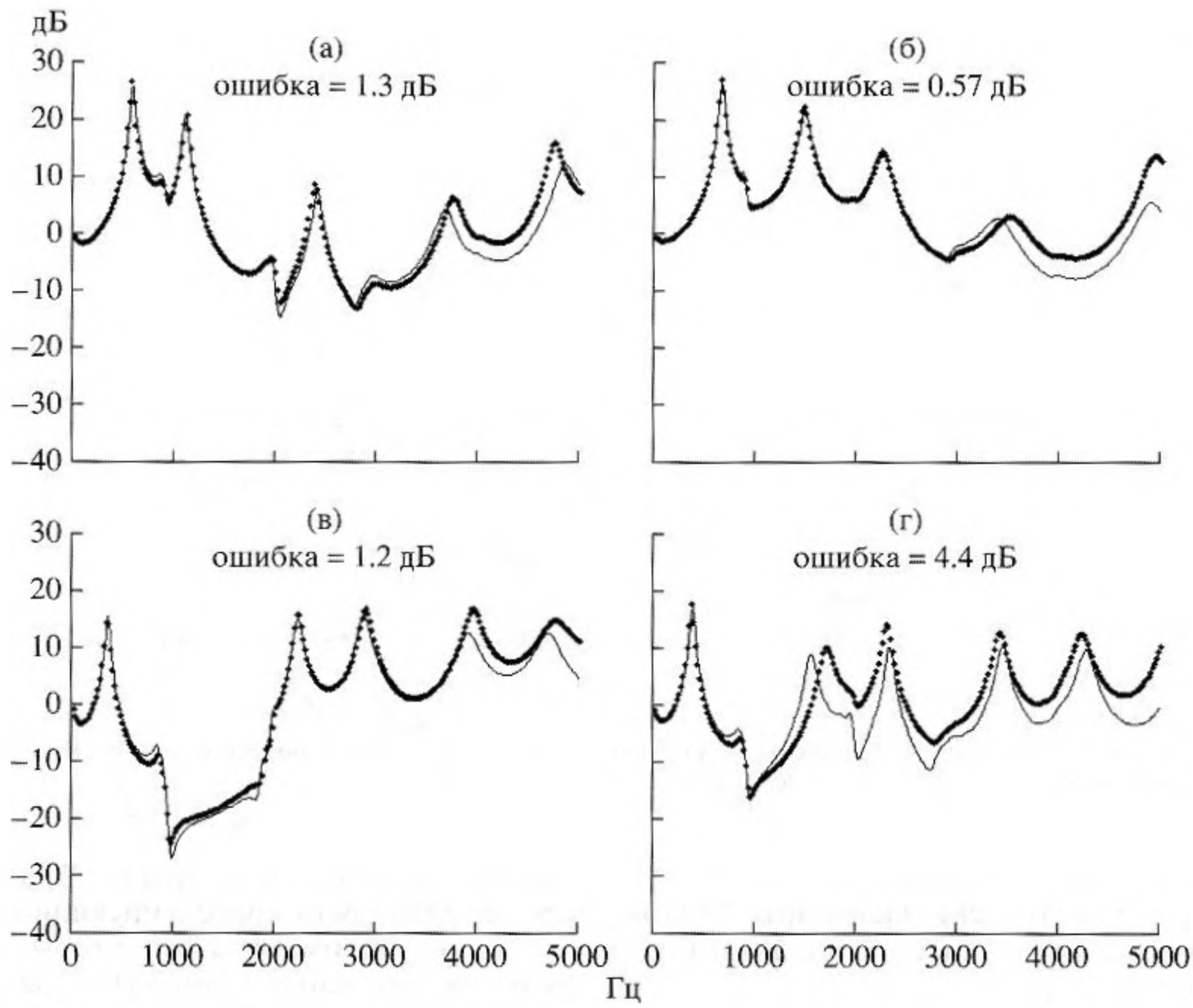

Рис. 4. Огибающие передаточных функций для первого диктора, назальное разветвление (сплошная линия – коническая аппроксимация, пунктир – цилиндрическая аппроксимация).

/u/ в произнесении первого диктора, однако его величина снизилась на 1.5 дБ по сравнению с неразветвленным трактом. Для всех конфигураций увеличение количества секций при цилиндрической аппроксимации или уменьшение коэффициентов раствора рупоров при конической аппроксимации (для введения скачков площади на стыке рупоров) приводит к значительному снижению спектрального искажения и ошибок по резонансным и антирезонансным частотам.

Все эксперименты, описанные выше, были проведены на речевом тракте, аппроксимированном последовательностью 19 конических секций. Возникает вопрос, насколько изменятся акустические характеристики тракта при уменьшении количества рупоров. Для исследования этого вопроса необходимо располагать алгоритмом автоматической сегментации функции площади последовательностью функций вида (7). В эквивалентной постановке задача формулируется как проблема нахождения оптимального распределения узлов кусочно-линейной интерполяции функции радиуса $r(x) = \sqrt{S(x)/\pi}$. Нами был использован алгоритм, описанный в [2]. Сначала находятся все локальные максимумы и минимумы функции $r(x)$, которые соединяются прямыми линиями. Затем на каждом интервале между максимумом и минимумом или минимумом и максимумом вычисляется среднеквадратическая ошибка линейной аппроксимации $r(x)$. Если эта ошибка превышает некоторый порог, то данный интервал разбивается на подынтервалы, в каждом из которых снова осуществляется линейная аппроксимация. После того, как в каждом интервале ошибка линейной аппроксимации оказывается меньше порога, полученная кусочно-линейная функция пересчитывается в площадь $S(x)$.

В проведенных экспериментах порог ошибки линейной аппроксимации для функции эквивалентного радиуса $r(x)$ полагался равным 10%. В результате работы алгоритма общее количество конических рупоров сократилось в среднем вдвое (по сравнению с исходной конфигурацией, аппроксимированной последовательностью 19 конических секций). Ошибки аппроксимации исходной площади последовательностью функций вида (7) по абсолютному значению не превосходили 8%. Соответствующие огибающие оказались очень близки как по форме, так и по расположению спектральных пиков. Для обоих дикторов невязки между резонансными и антирезонансными (при наличии разветвления) частотами, вычислен-

ными по обеим схемам длинной линии, по абсолютному значению не превосходили 3.5%. Средняя величина спектрального искажения для обоих дикторов составила 1.5 дБ. Отсюда следует, что для получения точных акустических характеристик речевого тракта достаточно аппроксимировать его небольшим числом конических рупоров.

ЗАКЛЮЧЕНИЕ

Построена схема длинной линии для речевого тракта, аппроксимированного последовательностью конических рупоров, в предположении, что в каждом рупоре распространяются плоские волны. Схема обобщена на все случаи, представляющие интерес для акустической теории речеобразования, – наличие потерь на вязкое трение и теплопроводность, податливость стенок речевого тракта, а также возможное наличие разветвляющих каналов. Схема протестирована на площадях поперечных сечений речевого тракта, измеренных с помощью магнитно-резонансной томографии на реальном дикторе. Результаты, полученные с помощью выведенных уравнений, сопоставлены с результатами, получаемыми с помощью стандартного метода длинной линии, основанного на цилиндрической аппроксимации речевого тракта. Показано, что аппроксимация речевого тракта небольшим (порядка десяти) числом конических рупоров и очень большим (более 100) числом цилиндрических труб дают сопоставимые оценки акустических характеристик тракта (резонансные и антирезонансные частоты, огибающие передаточных функций). Это свидетельствует о том, что построенная схема длинной линии является вычислительно проще, точнее и эффективнее, чем стандартная схема. В настоящее время нами исследуются вопросы использования построенной схемы длинной линии в артикуляторном синтезаторе, а также обобщение этой схемы на случай сферических волн, распространяющихся в каждом коническом рупоре.

СПИСОК ЛИТЕРАТУРЫ

1. *Сорокин В.Н.* Теория речеобразования. М.: Радио и связь. 1985. 312 с.
2. *Сорокин В.Н.* Синтез речи. М.: Связь. 1992. 392 с.
3. *Фланаган Дж.* Анализ, синтез и восприятие речи. М.: Связь. 1968. 392 с.
4. *Allen D., Strong W.* A Model for Synthesis of Natural Sounding Vowels // J. Acoust. Soc. Amer. 1985. V. 78. P. 58–69.
5. *Sondhi M.M., Schroeter J.* A Hybrid Time-Frequency Domain Articulatory Speech Synthesizer // IEEE Trans. Acoust., Speech, Signal Process., ASSP-35. 1987. P. 955–967.
6. *Zhang Zh., Espy-Wilson C.* A Vocal-Tract Model of American English/l/ // J. Acoust. Soc. Amer. 2004. V. 115. P. 1274–1280.
7. *Макаров И.С., Сорокин В.Н.* Резонансы разветвленного речевого тракта с податливыми стенками // Акуст. журн. 2004. Т. 50. № 3. С. 389–396.
8. *El-Masri S., Pelorson X., Saguet P., and Badin P.* Development of the Transmission Line Matrix Method in Acoustics Applications to Higher Modes in the Vocal Tract and Other Complex Ducts // Int. J. Numer. Model. 1998. V. 11. P. 133–151.
9. *Story B.* Technique for "Tuning" Vocal Tract Area Functions Based on Acoustic Sensitivity Functions // J. Acoust. Soc. Amer. 2006. V. 119. P. 715–718.
10. *Badin P., Fant G.* Notes on Vocal Tract Computation // STL-QPSR. 1984. P. 53–108.
11. *Causse R., Kergomard J., and Lurton X.* Input Impedance of Brass Musical Instruments – Comparision between Experiment and Numerical Models // J. Acoust. Soc. Amer. 1984. V. 75. P. 241–254.
12. *Adachi S., Yamada M.* An Acoustical Study of Sound Production in Biphonic Singing, Xoomij // J. Acoust. Soc. Amer. 1999. V. 105. P. 2920–2932.
13. *Adachi S., Takemoto H., Kitamura T., Mokhtari P., and Honda K.* Vocal Tract Length Perturbation and its Application to Male-Female Vocal Tract Shape Conversion // J. Acoust. Soc. Amer. 2007. V. 121. P. 3874–3885.
14. *Зайцев В.Ф., Полянин А.Д.* Справочник по линейным обыкновенным дифференциальным уравнениям. М.: Издательство "Факториал". 1997. 304 с.
15. *Sondhi M.M.* Model for Wave Propagation in a Lossy Vocal Tract // J. Acoust. Soc. Amer. 1974. V. 55. P. 1070–1075.
16. *Sondhi M.M., Resnick J.* The Inverse Problem for the Vocal Tract: Numerical Methods, Acoustical Experiments, and Speech Synthesis // J. Acoust. Soc. Amer. 1983. V. 73. P. 985–1002.
17. *Chu W.* Speech Coding Algorithms: Foundation and Evoluation of Standartized Coders. A John Wiley and Sons, Inc. 2003. P. 578.
18. *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.
19. *Fant G., Bavegard M.* Parametric Model of VT Area Functions: Vowels and Consonants // TMH-QPSR. 1997. V. 1. P. 1–20.
20. *Kewley-Port D., Watson Ch.* Formant-Frequency Discrimination for Isolated English Vowels // J. Acoust. Soc. Amer. 1994. V. 95. P. 485–496.
21. *Hawks J.* Difference Limens for Formant Patterns of Vowel Sounds // J. Acoust. Soc. Amer. 1994. V. 95. P. 1074–1084.
22. *Фант Г.* Акустическая теория речеобразования. М.: Наука. 1964. 284 с.
23. *Dang J., Shadle Ch., Kawanishi Y., Honda K., and Suziki H.* An Experimental Study of the Open End Correction Coefficient for Side Branches within an Acoustical Tube // J. Acoust. Soc. Amer. 1998. V. 104. P. 1075–1084.
24. *Ржевкин С.Н.* Курс лекций по теории звука. М: МГУ. 1960. 335 с.