АКУСТИЧЕСКИЙ ЖУРНАЛ, 2004, том 50, № 3, с. 389–396

БИОЛОГИЧЕСКАЯ АКУСТИКА

УДК 612.85

РЕЗОНАНСЫ РАЗВЕТВЛЕННОГО РЕЧЕВОГО ТРАКТА С ПОДАТЛИВЫМИ СТЕНКАМИ

© 2004 г. И. С. Макаров, В. Н. Сорокин

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

Поступила в редакцию 14.01.2003 г.

Вычисление резонансных частот по экспериментально измеренным площадям поперечных сечений речевого тракта в предположении абсолютной жесткости стенок приводит к существенному расхождению с измеренными резонансными частотами. Податливость стенок влияет на первый резонанс и практически не затрагивает высшие резонансы. Наличие разветвлений в тракте на уровне гортани влияет на второй и третий резонансы в большей степени, чем на первый. Параметры импеданса стенок (потери, масса, упругость), а также длина и площадь поперечного сечения разветвлений на уровне гортани определяются путем минимизации ошибки в смысле наименьших квадратов между измеренными и вычисленными резонансными частотами. Показано, что точность вычисленных частот с учетом податливости и наличия разветвлений в тракте оказывается в пределах точности оценки формант.

При артикуляторном синтезе речи часто возникает задача вычисления резонансных частот речевого тракта по заданной функции его площади поперечного сечения [1]. При решении обратной задачи относительно площади поперечного сечения или формы речевого тракта для гласных звуков, входные данные представляются в виде измеренных формантных частот [2]. Резонансные частоты речевого тракта с податливыми стенками обычно определяются как решение спектральной задачи с краевыми условиями для волнового уравнения

$$\frac{\partial}{\partial x}\left[S(x)\frac{\partial p}{\partial x}\right] = \frac{S(x)}{c_0^2(x)}\frac{\partial^2 p}{\partial t^2}, \qquad (1)$$

где $S(x)$ – площадь поперечного сечения речевого тракта, x – координата вдоль средней линии тракта, $p(x, t)$ – звуковое давление, $c_0(x)$ – скорость распространения звуковых волн в тракте. Это уравнение справедливо в диапазоне частот ниже 4.5 кГц для неизменной во времени площади $S(x)$ [3].

До появления методов непосредственного измерения площади поперечного сечения речевого тракта рассматривались различные методы определения резонансных частот по заданной площади поперечного сечения [4, 5]. В работах [6–8] было показано несовпадение между измеренными и вычисленными резонансами речевого тракта при измеренных методом магнитно-резонансной томографии (MRI) площадях тракта.

Известно, что импеданс стенок тракта инерционного или упругого типа может повышать или понижать частоту первого резонанса [5]. В нижней части речевого тракта имеются грушевидные разветвления (называемые piriform fossa или sinus piriformis), которые представляют собой две полости конической формы, соединяющиеся с речевым трактом по обеим боковым сторонам глотки приблизительно на уровне входа в пищевод (рис. 1). Грушевидные области могут оказать существенное влияние на резонансные частоты речевого тракта [7, 9]. Однако, данные разных работ противоречат друг другу. Так, в [9] сообщается, что наибольшее влияние грушевидные области имеют на первый резонанс (частота первого резонанса может понижаться на 6–10%), влияние же на высшие резонансы сравнительно мало (частоты высших резонансов могут снижаться на 1.2–3%). В работе [7], напротив, отмечалось, что грушевидные области в наибольшей степени влияют на частоты второго и третьего резонансов, и в меньшей степени – на частоту первого резонанса.

Таким образом, оказывается необходимым исследование факторов, влияющих на резонансные частоты речевого тракта, т.е. создание более реалистической акустической модели.

В настоящее время основным экспериментальным методом измерения площадей поперечных сечений тракта является магнитно-резонансная томография (MRI) [7]. MRI – это метод томографического отображения, служащий для послойных магнитно-резонансных изображений человеческого тела. При проведении сканирования к голове диктора приставляется набор пластин, расположенных поперек речевого тракта. Диктор располагается внутри магнита, который создает магнитное поле. Магнитное поле необходимо для поворота спинов водорода в тканях ре-

Рис. 1 Речевой тракт: вид сзади.

чевого тракта на 90° и 180°. Внутри магнита расположены радиочастотные катушки, которые возбуждают высокочастотными радиоимпульсами спины водорода в каждом слое, а также регистрируют отраженный ядерномагнитный сигнал. Этот сигнал подвергается двухмерному преобразованию Фурье, в результате чего получается магнитно-резонансное изображение каждого слоя [8]. Во время сканирования, которое обычно длится 5–15 с, диктор должен сохранять постоянной артикуляцию звука. Применение магнитно-резонансной томографии, в отличие от ранее использовавшихся рентгеновских измерений, не накладывает ограничения на время эксперимента с целью предотвращения недопустимого облучения.

Из-за сложностей, связанных с проведением магнитно-резонансных исследований, в настоящее время существует весьма небольшой набор экспериментальных данных о площадях поперечного сечения речевого тракта. Как правило, эти данные описывают стационарные гласные звуки.

Одним из основных методов оценивания резонансных частот является линейное предсказание. В работе [10] было показано, что систематическая ошибка определения частот формант, связанная с частотой основного тона, составляет около 10% от частоты основного тона. Систематические ошибки, связанные с выбором порядка модели, а также с методом нахождения полюсов и порядковым номером полюса, оказываются в пределах 10–80 Гц. Таким образом, оценка точности формант для метода линейного предсказания находится в пределах 8–10%.

Данные о площадях поперечного сечения гласных, измеренных с помощью магнитно-резонансной томографии, а также первые три резонансные частоты для каждого гласного были взяты из работ [6, 7]. Использовались 2 набора функций площадей. Каждый набор включал 10 функций площадей тракта во время артикуляции 10 английских гласных звуков /i/ (heat), /ι/ (hit), /ε/ (head), /æ/ (hat), /Λ/ (ton), /α/ (hot), /⊃/ (paw), /o/ (hoe), /∪/(hood), /u/(who). Первый набор данных описывает речевой тракт диктора-мужчины, а второй набор данных – женщины. При проведении магнитно-резонансной томографии использовался набор из 24 пластин толщиной 5 мм. Расстояние между пластинами было равно 5 мм.

Резонансные частоты в этих работах вычислялись в окне Хемминга длительностью 25 мс с помощью автокорреляционного метода линейного предсказания с порядком модели равным 16. Полюса передаточной функции вычислялись методом нахождения пиков в спектре Фурье с последующей параболической интерполяцией. Наряду с измеренными резонансными частотами, в работах [6, 7] по измеренным площадям поперечных сечений вычислялись резонансные частоты с помощью модифицированной схемы бегущих волн [11, 12]. В этой схеме учитывалась податливость стенок и импеданс излучения через губы. Разветвления в тракте в [6] не учитывались, а в [7] учитывались как дополнительная площадь в области глотки.

В теории речеобразования существуют различные модели вычисления резонансных частот по площадям поперечных сечений [1]. В проведенном исследовании в качестве акустической модели использовалась схема длинных линий, которая позволяет учесть податливость стенок, а также разветвления в тракте. В схеме длинных линий речевой тракт разбивается на N секций постоянного поперечного сечения. Давление P_l и объемная скорость U_l у губ связаны с давлением P_g и объемной скоростью U_g у голосовой щели соотношением (2):

$$\begin{pmatrix} P_g \\ U_g \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} P_l \\ U_l \end{pmatrix}, \tag{2}$$

где

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \prod_{i=1}^{N} \begin{pmatrix} \mathrm{ch}(\gamma l_i) & \dfrac{\rho \bar{c}_0}{S(i)} \mathrm{sh}(\gamma l_i) \\ \dfrac{S(i)}{\rho \bar{c}_0} \mathrm{sh}(\gamma l_i) & \mathrm{ch}(\gamma l_i) \end{pmatrix}. \tag{3}$$

Здесь N – общее количество секций, $S(i)$ – площадь поперечного сечения i-й секции, l_i – длина i-й секции, $\gamma = \beta + jk$, β – коэффициент затухания, $k = \omega/\bar{c}_0$ – волновое число, ω – круговая частота (рад/с), ρ – плотность воздуха (0.00114 г/см^3), $\bar{c}_0$ – скорость звука в трубе с жесткими стенками (35000 см/с), j – мнимая единица. Передаточная

Таблица 1. Резонансные частоты тракта с жесткими стенками без пириформических областей, диктор-мужчина (Titze 1996)

Гласный	F1 (изм)	F1 (выч)	Ошибка (%)	F2 (изм)	F2 (выч)	Ошибка (%)	F3 (изм)	F3 (выч)	Ошибка (%)
/i/ heat	333	226.5	–31.9	2332	2575	10.4	2986	3652	22.3
/ι/ hit	518	487.6	–5.9	2004	2141	6.9	2605	2729	4.8
/ε/ head	624	646.9	3.7	1853	2064	11.4	2475	2665	7.8
/æ/ hat	692	765.9	10.7	1873	1952	4.2	2463	2665	8.2
/Λ/ ton	707	753.8	6.7	1161	1363	17.4	2591	2679	3.4
/α/ hot	754	808	7.2	1195	1209	1.2	2685	2895	7.8
/⊃/ paw	654	644.9	–1.4	944	1040	10.2	2739	2213	–19.2
/o/ hoe	541	382.8	–29.2	1045	926	–11.4	2568	2721	5.9
/∪/ hood	540	395	–26.8	922	893	–3.1	2584	2547	–1.4
/u/ who	389	261.5	–32.8	987	1206	22.2	2299	2480	7.9

Таблица 2. Резонансные частоты тракта с жесткими стенками без пириформических областей, диктор-женщина (Titze 1998)

Гласный	F1 (изм)	F1 (выч)	Ошибка (%)	F2 (изм)	F2 (выч)	Ошибка (%)	F3 (изм)	F3 (выч)	Ошибка (%)
/i/ heat	489	316.6	–35.3	2387	2866	20.0	3526	3262	–7.5
/ι/ hit	649	714	10.0	2079	2328	11.9	2799	3511	25.4
/ε/ head	799	856.7	7.2	2112	2312	9.5	2874	3394	18.0
/æ/ hat	837	1016	21.4	2028	2536	25.0	2814	4042	43.6
/Λ/ ton	759	903.6	19.0	1360	1604	17.9	2756	3577	29.8
/α/ hot	961	828.4	–13.8	1488	1644	10.5	2779	3058	10.0
/⊃/ paw	817	819.9	1.0	1264	1369	8.3	2574	2963	15.1
/o/ hoe	706	777.6	10.2	1157	1588	37.3	3030	3281	8.3
/∪/ hood	675	920.9	36.4	1122	1616	44.0	3157	3098	–1.9
/u/ who	533	458.5	–13.9	1152	1233	7.0	3122	3162	1.3

функция речевого тракта определяется как отношение объемной скорости у губ к объемной скорости у голосовой щели, т.е. $H(\gamma) = U_l/U_g$. Входной импеданс со стороны голосовой щели есть отношение давления у голосовой щели к объемной скорости у голосовой щели, т.е. $Z_g = P_g/U_g$. Из формул (2)–(3) следует, что

$$H(\gamma) = \frac{1}{CZ_l + D}, \tag{4}$$

$$Z_g = \frac{AZ_l + B}{CZ_l + D}. \tag{5}$$

Здесь Z_l – импеданс излучения у губ, определяемый как $Z_l = \rho\frac{(2\pi F)^2}{2c_0} + j\frac{16F\rho}{3R_0}$ [13]. В этой формуле R_0 – эквивалентный радиус губного отверстия, F – частота.

Из (4), (5) следует, что полюса передаточной функции и входного импеданса совпадают, так что для определения резонансных частот достаточно по методу длинных линий вычислить либо передаточную функцию, либо входной импеданс. В проведенном исследовании всюду вычислялся входной импеданс.

Результаты вычислений резонансных частот представлены в таблицах 1 и 2. Из этих таблиц видно, что резонансные частоты, вычисленные с помощью алгоритма длинных линий по экспериментально измеренным площадям поперечных сечений в предположении жестких стенок, во многих случаях весьма сильно отличались от измеренных резонансных частот. При этом ошибка вычислений зачастую превышает ошибку точности оценки формант. Для первого резонанса это проявляется на гласных /i/ (heat), /æ/ (hat), /o/ (hoe), /∪/ (hood), /u/ (who) (для обоих дикторов), а также на гласных /ι/ (hit), /Λ/ (ton), /α/ (hot) – для диктора-женщины. Для второго резонанса наибольшие ошибки наблюдаются на гласных в словах гласных /i/ (heat), /o/ (hoe), /Λ/ (ton) (для обоих дикторов), /ε/ (head), /⊃/ (paw), /u/ (who) – для диктора-мужчины и для /ι/ (hit), /æ/ (hat), /∪/ (hood) – для диктора-женщины. Для третьего резонанса наибольшие ошибки наблюдаются на гласном в слове /⊃/ (paw) для обоих дикторов, на гласном в слове /i/ (heat) для диктора-мужчины и на гласных в словах /ι/ (hit), /ε/ (head), /æ/ (hat), /Λ/ (ton) – для диктора-женщины. Частоты, полученные с помощью этого алгоритма, были сопоставлены с резонансными частотами, вычисленными другими методами, а именно, конечно-разностной схемой,

описанной применительно к речевому сигналу в [1], а также схемой определения полюсов передаточной функции, вычисленной с помощью метода бегущих волн [14]. Значения резонансных частот, вычисленных с помощью этих методов, а также резонансные частоты, вычисленные в [6, 7], оказались близкими. Таким образом, наблюдаемые несовпадения между вычисленными и измеренными резонансными частотами объясняются не особенностями алгоритма, а моделью речевого тракта.

Как видно из табл. 1, 2, второй и третий вычисленные резонансы обычно оказываются выше второго и третьего измеренных резонансов. Что касается первого вычисленного резонанса, то он для диктора-мужчины обычно оказывается либо выше первого измеренного резонанса, но в пределах точности оценки формант, либо значительно ниже первого измеренного резонанса. Для женщины-диктора первый вычисленный резонанс может быть как значительно ниже, так и значительно выше первого измеренного резонанса.

Поскольку на первый резонанс в наибольшей степени влияет податливость стенок, была исследована роль параметров импеданса стенок. Податливость стенок речевого тракта приводит к тому, что уменьшение столба воздуха в тракте под воздействием давления обуславливается не только сжимаемостью самой среды, но и смещением стенок. В [15] было показано, что скорость звука в трубе с податливыми стенками определяется как

$$c_0^2 = \frac{\overline{c_0^2}}{1 - \dfrac{2\rho\overline{c_0^2}}{j\omega Z_\omega a}}, \qquad (6)$$

где Z_ω – импеданс стенок, a – характерный поперечный геометрический размер трубы, $\bar{c}_0$ – скорость звука в трубе с абсолютно жесткими стенками. Область применения формулы (6) ограничивается толщиной стенок трубы, так как если она сравнима с длиной радиальной волны, то свойства стенок не исчерпываются их импедансом, а зависят от распределения звукового давления и скорости частиц вдоль трубы [16]. Стенки речевого тракта могут считаться тонкими по сравнению с длиной волны радиальных колебаний, и (6) справедливо для оценки акустических процессов в речевом тракте. В общем виде импеданс стенок Z_ω может быть записан как

$$Z_\omega = r + j\left(\frac{c}{\omega} - \omega m\right), \qquad (7)$$

где r – потери на стенках, m – масса стенок, c – упругость стенок.

Как видно из (6), податливость стенок влияет на скорость звука в области относительно низких частот [5, 17]. Непосредственное использование соотношения (6) в методе длинных линий для вычисления резонансных частот связано с рядом сложностей. В случае произвольной зависимости геометрического параметра a от координаты x получается система нелинейных уравнений, решение которой затруднительно. Поэтому была выбрана предложенная в [18] модифицированная схема длинных линий, которая учитывает параметры импеданса стенок. В этой схеме соотношение (2) остается в силе, а соотношение (3) заменяется соотношением

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \prod_{i=1}^{N} \begin{pmatrix} \operatorname{ch}\left(\dfrac{\sigma l_i}{\bar{c}_0}\right) & \dfrac{\rho\bar{c}_0}{S(i)}\varphi\operatorname{sh}\left(\dfrac{\sigma l_i}{\bar{c}_0}\right) \\ \dfrac{S(i)}{\rho\bar{c}_0\varphi}\operatorname{sh}\left(\dfrac{\sigma l_i}{\bar{c}_0}\right) & \operatorname{ch}\left(\dfrac{\sigma l_i}{\bar{c}_0}\right) \end{pmatrix}. \qquad (8)$$

Здесь l_i – длина i-й секции, $\varphi = -\sqrt{\omega}/\sqrt{\theta + j\omega}$, ω – круговая частота, $\sigma = \varphi(\theta + j\omega)$, $\theta = \dfrac{j2.6\bar{c}_0\pi^2\omega}{j\omega(1+\chi)+\delta}$, $\chi = \dfrac{r}{m}$, $\delta = \dfrac{1}{mc}$.

Ткани стенок речевого тракта имеют различные характеристики. Зубы можно считать абсолютно жесткой поверхностью. Язык, губы, щеки и мягкое небо более или менее податливы, а твердое небо имеет костную основу, на которой находится сравнительно тонкий слой мягкой ткани. Поэтому естественно считать, что параметры импеданса стенок – потери, масса и упругость – в общем случае являются функцией координаты вдоль тракта.

Параметры импеданса стенок для каждого диктора находились путем минимизации ошибки в смысле наименьших квадратов разностей между вычисленной и измеренной первой резонансными частотами. Переменные по тракту потери, масса и упругость представлялись в виде первых восьми членов ряда Фурье:

$$r(x) = R_0 + \sum_{\xi=1}^{4} r_{1\xi}\sin\left(\frac{\xi\pi x}{l}\right) + \sum_{k=1}^{4} r_{2\xi}\cos\left(\frac{\xi\pi x}{l}\right),$$

$$m(x) = M_0 + \sum_{\xi=1}^{4} m_{1\xi}\sin\left(\frac{\xi\pi x}{l}\right) + \sum_{k=1}^{4} m_{2\xi}\cos\left(\frac{\xi\pi x}{l}\right),$$

$$c(x) = C_0 + \sum_{\xi=1}^{4} c_{1\xi}\sin\left(\frac{\xi\pi x}{l}\right) + \sum_{\xi=1}^{4} c_{2\xi}\cos\left(\frac{\xi\pi x}{l}\right).$$

Здесь l – длина тракта. В процессе оптимизации параметров r, c, m для каждого диктора по всем гласным звукам минимизировался функционал

$$\Phi(z) = \sum_{n=1}^{T} [F_m^{(1n)} - F_c^{(1n)}(z)]^2 \qquad (9)$$

Таблица 3. 1-й резонанс без учета и с учетом податливости стенок, диктор-мужчина (Titze 1996), потери = 542.4 г/с см^2, масса = 2.4 г/см^2, упругость = 1.5×10^3 Па/см

Гласный	F1 (изм)	F1 (выч. без податл)	Ошибка (%)	F1 (выч. с пост. податл)	Ошибка (%)	F1 (выч. с перем. податл)	Ошибка (%)
/i/ heat	333	226.5	–31.9	375.9	12.9	375.3	12.7
/ɪ/ hit	518	487.6	–5.9	551.6	6.5	549.1	6
/ɛ/ head	624	646.9	3.7	635	1.8	625.9	0.3
/æ/ hat	692	765.9	10.7	691	–1.1	680.2	–1.7
/ʌ/ ton	707	753.8	6.7	685.7	–3.2	684.4	–3.2
/ɑ/ hot	754	808	7.2	710	–5.8	708.8	–6
/ɔ/ paw	654	644.9	–1.4	634	–3.1	630.7	–3.56
/o/ hoe	541	382.8	–29.2	488.8	–9.6	492.9	–8.9
/ʊ/ hood	540	395	–26.8	497	–8.2	498.9	–7.6
/u/ who	389	261.5	–32.8	403.9	3.8	399.9	2.8

с ограничениями $R = \{r, r1 \le \min(r); \max(r) \le r2\}$, $M = \{m, m1 \le \min(m); \max(m) \le m2\}$, $C = \{c, c1 \le \le \min(c); \max(c) \le c2\}$. Здесь z – вектор с компонентами (r, m, c), $T = 10$ – количество гласных звуков, использованных в экспериментах, $F_m^{(1n)}$ – первая измеренная резонансная частота для n-го звука, $F_c^{(1n)}$ – первая вычисленная резонансная частота для n-го звука. По различным данным непосредственных измерений импеданса тканей речевого тракта [5] потери находятся в диапазоне 540–11000 г/с · см^2, массы – в диапазоне – 1.08–2.7 г/с · см^2, а упругость в диапазоне от 10^3 до 10^6 Па/см. Соответственно этим данным ограничения R, M, C были определены как $R = \{r, 540 \le \le r \le 11000\}$, $M = \{m, 1.08 \le m \le 2.7$ г/с · см$^2\}$, $C = = \{c, 10^3 \le c \le 10^6$ Па/см$\}$ [5]. Наряду с переменными параметрами стенок, исследовалось влияние эффективных (постоянных) параметров. В качестве минимизирующего функционала использовался функционал (9). Ограничения R, M, C определялись так же, как и в схеме с переменными по тракту параметрами импеданса стенок.

Из таблиц 3, 4 видно, что использование переменных по тракту параметров импеданса стенок практически не улучшает результаты по сравнению с постоянными вдоль тракта параметрами. В некоторых случаях (/ɪ/ (hit), /ʌ/ (ton), /ɑ/ (hot), /u/ (hood) для диктора-женщины), переменные по тракту параметры даже приводят к худшим результатам. Поэтому при вычислении первой резонансной частоты с учетом податливости достаточно положить параметры импеданса стенок постоянными по тракту.

Поскольку податливость стенок в наибольшей степени влияет на первый резонанс, была проведена проверка того, в какой степени учет разветвлений позволяет улучшить значения второй и третьей вычисленной резонансной частоты.

Резонансные частоты разветвленного тракта определялись по полюсам входного импеданса следующим образом: сначала по формулам (5) и (8) вычислялся входной импеданс Z_m в область от точки разветвления до губ; по тем же формулам вычислялся входной импеданс Z_p в грушевидную область со стороны речевого тракта (в предположении, что с другой стороны грушевидная область оперта на абсолютно жесткую стенку). Затем входной импеданс Z_{in} в область от точки разветвления до губ пересчитывался по формуле

$$Z_{in} = \frac{Z_m Z_p}{Z_m + Z_p}. \tag{10}$$

Наконец, входной импеданс в тракт определялся как

$$Z_g = \frac{AZ_{in} + B}{CZ_{in} + D}. \tag{11}$$

Оба грушевидных разветвления можно моделировать двумя цилиндрическими трубами постоянного поперечного сечения, шунтирующими речевой тракт на расстоянии примерно 2 см от голосовой щели. С вычислительной точки зрения предпочтительнее моделировать оба разветвления одной эквивалентной трубой. С этой целью были построены две модели – одна, учитывающая (помимо речевого тракта) два разветвления (трехтрубная модель), и вторая с одним разветвлением (двухтрубная модель). Трехтрубная модель зависит от четырех параметров – двух площадей поперечных сечений и двух длин труб. Двухтрубная модель определяется двумя параметрами – площадью поперечного сечения и длиной трубы. Параметры каждой модели были вычислены для каждого диктора путем минимизации ошибки в смысле наименьших квадратов между тремя вычисленными и измеренными резонансными частотами. В каждой модели стенки тракта принимались абсолютно жесткими. Оказалось, что ошибки между частотами, вычисленными на основе трехтрубной и двухтрубной модели, не

Таблица 4. 1-й резонанс без учета и с учетом податливости стенок, диктор-женщина (Titze 1998), потери = 541.9 г/с см^2, масса = 2.59 г/см^2, упругость = 1.4×10^5 Па/см

Гласный	F1 (изм)	F1 (выч. без податл)	Ошибка (%)	F1 (выч. с пост. податл)	Ошибка (%)	F1 (выч. с перем. податл)	Ошибка (%)
/i/ heat	489	316.6	–35.3	461.8	–5.6	481.2	–1.6
/ɪ/ hit	649	714	10.0	693.7	6.8	699.6	7.8
/ɛ/ head	799	856.7	7.2	759.6	–4.9	783.8	–1.9
/æ/ hat	837	1016	21.4	827.5	–1.2	826.9	–1.2
/ʌ/ ton	759	903.6	19.0	780	2.7	794.7	4.7
/ɑ/ hot	961	828.4	–13.8	747	–22.3	739.9	–23.0
/ɔ/ paw	817	819.9	1.0	743.2	–9.0	776.2	–5.0
/o/ hoe	706	777.6	10.2	723.8	2.5	716.6	1.5
/ʊ/ hood	675	920.9	36.4	787.6	16.6	789.8	17.0
/u/ who	533	458.5	–13.9	555.7	4.2	534.1	0.2

превышают 0.8%. Поэтому в дальнейшем была использована только двухтрубная модель.

Затем была предпринята попытка определить совместно параметры импеданса стенок и параметры разветвляющей трубы (длина и площадь поперечного сечения). Параметры импеданса стенок и шунтирующей трубы для каждого диктора находились путем минимизации функционала (12)

$$\Phi(z_1) = \sum_{n=1}^{N} (F_m^{(1n)} - F_c^{(1n)}(z_1))^2 + \\ + (F_m^{(2n)} - F_c^{(2n)}(z_1))^2 + (F_m^{(3n)} - F_c^{(3n)}(z_1))^2 \quad (12)$$

с ограничениями $R = \{r, r1 \le r \le r2\}$, $M = \{m, m1 \le \le m \le m2\}$, $C = \{c, c1 \le c \le c2\}$, $S_p = \{s_p, s_p1 \le s_p \le s_p2\}$, $L_p = \{l_p, l_p1 \le l_p \le l_p2\}$.

Здесь z_1 – вектор с компонентами (z, m, c, s_p, l_p), s_p – площадь поперечного сечения разветвляющей трубы, l_p – длина этой трубы, $F_m^{(\theta n)}$ – измеренная резонансная частота, $F_c^{(\theta n)}$ – вычисленная резонансная частота, n – порядковый номер гласного, θ – порядковый номер резонансной частоты. Ограничения R, M, C были определены выше. По данным [7, 9], длина грушевидной области может варьироваться от 1 до 3 см, а площадь оказывается в пределах от 0.05 до 1.5 см^2. Поэтому ограничения S_p, L_p были определены как $S_p = \{s_p, 0.05 \text{ см}^2 \le \le s_p \le 1.5 \text{ см}^2\}$, $L_p = \{l_p, 1 \text{ см} \le l_p \le 3 \text{ см}\}$.

Результаты вычислений приведены в таблицах 1, 2. Из таблиц видно, что одновременный учет податливости стенок и наличия разветвлений приводит к тому, что ошибки между вычисленными и измеренными резонансными частотами в большинстве случаев оказываются в пределах точности оценки формант. Иногда совместный учет податливости стенок и разветвлений приводит к некоторому ухудшению погрешности. Например, для диктора-мужчины это гласный /ɔ/ в слове paw, где ошибка по третьему резонансу с –19.2% увеличивается на 0.3%, однако такое возрастание ошибок представляется несущественным. Для диктора-женщины это гласный /ɑ/ в слове hot, где ошибка по первому резонансу с –13.8% увеличивается до – 21.3%, а также гласный /ʊ/ в слове hood, где ошибка по третьему резонансу с –1.9% увеличивается до –16.9%. В некоторых случаях учет разветвлений в тракте приводит к снижению ошибки по второму и третьему резонансу, однако достигаемый уровень ошибки все же существенно превосходит уровень точности оценки формант. Для диктора-женщины это гласный в слове hat, где ошибка по третьему резонансу с 43.6% снижается до 27.1%. В одном случае (гласный /u/ в слове who для диктора-мужчины) учет разветвлений в тракте никак не повлиял на второй резонанс (ошибка с 22.2% уменьшилась на 0.1%).

В дополнение к экспериментальным данным [6, 7], были использованы данные работы [8]. В этой работе с помощью магнитно-резонансной томографии были получены формы тракта при артикуляции четырех гласных (/ɑ/ (hot), /æ/ (hat), /i/ (heat), /u/ (who)) для двух дикторов. Экспериментальная методика этой работы критиковалась в [6]. Поэтому эти данные использовались как своего рода контрольный материал.

Для каждого гласного были вычислены резонансные частоты с учетом податливости стенок и разветвлений. Совместный учет податливости стенок и разветвлений в тракте снижает средние ошибки по всем резонансным частотам на 2–5%. Вместе с тем ошибка по второму резонансу для диктора-мужчины (20.2%) и ошибки по первому и второму резонансам для диктора-женщины (18.1% и 35.7% соответственно) все же превосходят точность оценки формант. Эти случаи могут быть объяснены следующим образом. В настоящее время невозможно записывать акустический сигнал синхронно со сканированием формы тракта. Кроме того, при проведении сканирования тракта с помощью магнитно-резонансной томографии диктору приходится сохранять постоян-

Таблица 5. Резонансные частоты с учетом податливости стенок и пириформических областей, диктор-мужчина (Titze 1996), потери = 542.7 г/с см^2, масса = 2.43 г/см^2, упругость = 1.5×10^3 Па/см, длина пириформического разветвления = 2.1 см, площадь пириформического разветвления = 0.073 см^2

Гласный	F1 (изм)	F1 (выч)	Ошибка (%)	F2 (изм)	F2 (выч)	Ошибка (%)	F3 (изм)	F3 (выч)	Ошибка (%)
/i/ heat	333	377	13.1	2332	2506	7.4	2986	3334	11.6
/ι/ hit	518	553	6.7	2004	2095	4.5	2605	2657	2.0
/ε/ head	624	636	1.9	1853	1996	7.7	2475	2623	5.9
/æ/ hat	692	693	0.1	1873	1885	1.0	2463	2544	3.3
/Λ/ ton	707	685	–3.1	1161	1333	14.8	2591	2590	–0.1
/α/ hot	754	704	–6.6	1195	1198	–0.1	2685	2872	6.9
/⊃/ paw	654	634	–3.1	944	1021	8.1	2739	2205	–19.5
/o/ hoe	541	489	–9.4	1045	917	–12.2	2568	2695	4.9
/∪/ hood	540	497	–7.9	922	889	–3.5	2584	2459	–4.8
/u/ who	389	405	4.1	987	1206	22.1	2299	2448	6.5

Таблица 6. Резонансные частоты с учетом податливости стенок и пириформических областей, диктор-женщина (Titze 1998), потери = 548.1 г/с см^2, масса = 2.7 г/см^2, упругость = 1.5×10^3 Па/см, длина пириформического разветвления = 2.0 см, площадь пириформического разветвления = 0.72 см^2

Гласный	F1 (изм)	F1 (выч)	Ошибка (%)	F2 (изм)	F2 (выч)	Ошибка (%)	F3 (изм)	F3 (выч)	Ошибка (%)
/i/ heat	489	481	–1.6	2387	2596	8.7	3526	3097	–12.2
/ι/ hit	649	703	8.3	2079	1930	–7.1	2799	3013	7.6
/ε/ head	799	753	–5.7	2112	1817	–13.9	2874	3017	4.9
/æ/ hat	837	846	1.0	2028	2107	3.9	2814	3577	27.1
/Λ/ ton	759	742	–2.2	1360	1391	2.3	2756	3133	13.7
/α/ hot	961	756	–21.3	1488	1363	–8.4	2779	2916	4.9
/⊃/ paw	817	754	–7.7	1264	1139	–9.9	2574	2853	10.8
/o/ hoe	706	737	4.5	1157	1256	8.6	3030	2959	–2.4
/∪/ hood	675	734	8.8	1122	1282	14.3	3157	2625	–16.9
/u/ who	533	576	8.1	1152	1195	3.8	3122	2752	–11.9

ной артикуляцию гласного звука в течение довольно длительного промежутка времени (около 5–15 мс). По мнению [7, 8], именно эти обстоятельства являются основным источником значительных несовпадений между измеренными и вычисленными резонансными частотами.

На результаты магнитно-резонансной томографии также влияет расположение пластин, используемых в процессе сканирования речевого тракта. Разные расположения пластин относительно средней линии тракта приводят к разным площадям поперечных сечений, что отражается на вычисленных резонансных частотах. В [8] значительные несовпадения между вычисленными и измеренными резонансными частотами (которые для гласного /u/ (who) достигали 70%) объяснялись также неудачным выбором расположения пластин.

Ошибки могут быть также связаны с тем, что разрешение магнитно-резонансной томографии недостаточно для сканирования областей тракта с малыми (менее 0.2 см^2) площадями поперечных сечений. В [6] было показано, что при разрешении пластин 0.0938 см/пиксель ошибка между измеренными и реальными площадями узких труб (с площадью сечения 0.2 см^2) составляет около 10%. При уменьшении площади сечения трубы ошибки возрастают. Поэтому с помощью магнитно-резонансной томографии трудно оценить геометрические размеры узких разветвлений в речевом тракте (например, грушевидные разветвления).

С помощью магнитно-резонансной томографии трудно восстановить точную форму зубов из-за малого содержания водорода в них [8]. Поэтому для получения площадей поперечных сечений в передней части тракта приходится проводить измерения на искусственном небе и зубных протезах, сделанных специально для данного диктора. Эта процедура также может быть источником ошибок в оценке площадей в области твердого неба.

Еще один возможный источник погрешности связан с тем, что в проведенном исследовании параметры грушевидных разветвлений предполагались постоянными. Однако экспериментальные исследования, проведенные с помощью электролучевой компьютерной томографии, показали [7], что в процессе артикуляции грушевидные области могут активно изменять свою длину и объем. Можно предположить, что управление длиной и площадью поперечного сечения разветвлений могло бы привести к лучшему совпадению изме-

ренных и вычисленных резонансных частот. Однако в настоящее время учесть переменные параметры разветвлений в артикуляторной модели невозможно, поскольку экспериментальных данных об активности этих областей еще недостаточно.

Еще один источник ошибок может быть связан с тем, что в проведенном исследовании импеданс голосовой щели полагался бесконечным. Вместе с тем в [5] показано, что частоты резонансных колебаний могут подвергаться изменениям на интервале открытой голосовой щели. Кроме того, известно [19], что примерно у 80% женщин и 20% мужчин задняя часть голосовой щели остается открытой даже на интервале схлопывания голосовых складок, так что связь между ротовой полостью и подсвязочной областью не прерывается. Этот фактор также может влиять на резонансные частоты.

Еще одним неучтенным фактором ошибок может быть назализация, т.е. опускание небной занавески и разветвление речевого тракта в носовую область. Назализация приводит к смещению резонансных частот. Однако ее влияние трудно отследить с помощью линейного предсказания.

Податливость стенок тракта и наличие разветвлений существенно влияет на резонансные частоты. Формантные частоты, вычисленные в предположении абсолютно жестких стенок, значительно отличаются от измеренных формантных частот. Учет податливости стенок приводит к снижению ошибки по первому резонансу. При этом параметры импеданса стенок – потери, масса и упругость – могут быть выбраны постоянными по тракту. Точность вычисления первых трех резонансов для тракта с податливыми стенками и грушевидными разветвлениями оказывается в пределах точности оценки формант по речевому сигналу.

Работа выполнена при поддержке РФФИ, грант 03-01-0016.

СПИСОК ЛИТЕРАТУРЫ

1. *Сорокин В.Н.* Синтез речи. М.: Связь, 1992. 392 с.
2. *Sorokin V.N., Leonov A.S., Trushkin A.V.* Estimation of stability and accuracy of inverse problem solution for the vocal tract // Speech Communication. 2000. V. 30. C. 55–74.
3. *Фланаган Дж.* Анализ, синтез и восприятие речи. М.: Связь, 1968. 392 с.
4. *Фант Г.* Акустическая теория речеобразования. М.: Наука, 1964. 284 с.
5. *Сорокин В.Н.* Теория речеобразования. М.: Радио и связь, 1985. 312 с.
6. *Titze I., Story B., Hoffman E.* 1996 Vocal tract area functions from magnetic resonance imaging // J. Acoust. Soc. Amer. V. 100. № 1. P. 537–554.
7. *Story B., Titze I., Hoffman E.* Vocal tract area functions for an adult female speaker based on volumetric imaging // J. Acoust. Soc. Amer. 1998. V. 104. № 1. P. 471–487.
8. *Baer T., Gore J.C., Gracco L.C., Nye P.W.* Analysis of vocal tract shape and dimensions using magnetic-resonance imaging: Vowels // J. Acoust. Soc. Amer. 1991. V. 90. P. 799–828.
9. *Dang J., Honda K.* Acoustic characteristics of the piriform fossa in models and humans // J. Acoust. Soc. Amer. 1997. V. 101. № 1. P. 456–465.
10. *Vallabha G.K., Tuller B.* Systematic errors in the formant analysis of steady-state vowels // Speech Communication. 2002. V. 38. P. 141–160.
11. *Liljencrants J.* Speech synthesis with a reflection-type line analog. DS Dissertation. Dept. of Speech Comm. and Music Acoust., Royal Inst. of Tech., Stockholm, Sweden, 1985. 395 c.
12. *Story B.* Physiologically-based speech simulation using an enhanced wave-reflection model of the vocal tract. Ph.D. Dissertation, University of Iowa, 1995. 412 c.
13. *Морз Ф.* Колебания и звук. М.-Л.: ГИТТЛ, 1949. 496 с.
14. *Рабинер Л., Шафер Р.* Цифровая обработка речевых сигналов. М.: Радио и связь, 1981. 495 с.
15. *Исакович М.А.* Общая акустика. М.: Наука, 1973. 495 с.
16. *Скучик Е.* Основы акустики. Т. 2. М.: ИЛ, 1959. 565 с.
17. *Sondhi M.M.* Model for wave propagation in a lossy vocal tract // J. Acoust. Soc. Amer. 1974. V. 55. № 5. P. 1070–1075.
18. *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.
19. *Klatt D. H.* Software for a cascade/parallel formant synthesizer // J. Acoust. Soc. Amer. 1979. V. 67. № 3. P. 971–995.

Resonances of a Branched Vocal Tract with Compliant Walls

I. S. Makarov and V. N. Sorokin

The calculation of the resonance frequencies from the experimental cross-sectional areas of the vocal tract under the assumption that its walls are perfectly rigid provides the values that noticeably differ from the measured resonance frequencies. The compliance of the walls affects the first resonance and almost does not affect the higher-order resonances. The presence of branching in the tract at the level of larynx affects the second and third resonances stronger than the first resonance. The parameters of the wall impedance (the loss, mass, and elasticity) and the length and cross-sectional area of the branchings are determined by minimizing the rms discrepancy between the measured and calculated resonance frequencies. The error in the frequency calculation with allowance for the wall compliance and branching in the tract proves to be within the accuracy of the formant estimation.