

Моделирование подсвязочной области в частотно-временном артикуляторном синтезаторе

Горбунов К.С.,

Макаров И.С.,

*кандидат технических наук,
старший научный сотрудник*

Статья посвящена моделированию влияния подсвязочной области на акустические характеристики речевого тракта в частотно-временных артикуляторных синтезаторах. Построенная модель включает в себя вычисление кратковременного спектра Фурье голосового источника, учёт взаимодействия передаточной функции речевого тракта и подсвязочных полостей в частотной области и дальнейший синтез речевого сигнала методом наложения с суммированием. Схема протестирована на результатах измерений динамики речевого тракта методом магнитно-резонансной томографии для ряда звукосочетаний американского английского языка.

1. Постановка проблемы

Подсвязочная область представляет из себя систему полостей, расположенных ниже голосовых складок, — трахею, бронхи и лёгкие. Её влияние на акустические и аэродинамические процессы в речевом тракте разнообразно. В частности, при раскрытии голосовой щели создаются условия для возникновения дополнительных резонансов и антирезонансов, обусловленных влиянием подсвязочной области [Сорокин, 1985]. При этом резонансы речевого тракта модулируются как по амплитуде, так и по частоте. В некоторых случаях возможна бифуркация частоты первого резонанса речевого тракта, наблюдаемая на сонограммах в виде скачка амплитуды 100–300 Гц [Stevens, 2000; Chi, Sonderegger, 2007].

Таким образом, в результате взаимодействия подсвязочной области и речевого тракта в акустическом сигнале на интервале открытой голосовой щели наблюдаются спектральные компоненты, которые отсутствуют при закрытой голосовой щели. Влияние этих компонент на восприятие речи представляет значительный интерес.

По мнению [Сорокин, 1992], они определяют тембровые характеристики голоса и создают индивидуальные особенности звучания в зависимости от геометрических размеров трахеи, бронхов и лёгких. Согласно [Assmann, Katz, 2005], дополнительные резонансы и антирезонансы, порождённые подсвязочной областью, определяют не только натуральность, но и разборчивость стационарных гласных. Параметры трахеи и лёгких, вероятно, играют важную роль в создании ряда акустических эффектов в пении [Морозов, 2002].

Несмотря на отмеченные публикации, взаимодействие подсвязочной области и речевого тракта продолжает оставаться малоизученной проблемой. Нам, например, неизвестны работы, посвящённые моделированию этого взаимодействия для оценки его роли в восприятии речи.

На наш взгляд, это обстоятельство связано с трудностями моделирования подсвязочной области в существующих типах артикуляторных синтезаторов. В самом деле, во временных артикуляторных синтезаторах — то есть в синтезаторах, вычисляющих речевую волну во временной области либо методом бегущих волн, либо методом конечных разностей [Сорокин, 1992], — влияние подсвязочной области не может быть адекватно учтено из-за наличия нелинейных функций частоты в переносном импедансе трахеи. В частотно-временных синтезаторах источники возбуждения звука вычисляются во временной области, а характеристики тракта — в частотной [Сорокин, 1992]. При этом предполагается, что характеристики тракта (его передаточная функция и переносной акустический импеданс) изменяются во времени гораздо медленнее, чем площадь голосовой щели. По этой причине быстрые изменения спектральных характеристик сигнала, происходящие внутри одного периода основного тона и порождённые взаимодействием речевого тракта и подсвязочной области, принципиально не могут быть реализованы данным типом синтезатора без существенной переделки его общей схемы.

Основная задача данной работы состоит в построении такой схемы частотно-временного артикуляторного синтезатора речи, которая позволила бы моделировать влияние подсвязочной области на акустические характеристики речевого тракта. Основная идея построения заключается в вычислении кратковременного спектра Фурье голосового источника в скользящем окне анализа, учёте взаимодействия речевого тракта и подсвязочных полостей в частотной области и дальнейшем синтезе речевого сигнала методом наложения с суммированием. Предлагаемая схема напоминает процедуру ресинтеза сигнала в вокоде STRAIGHT [Kawahara et al., 1999].

Структура работы выглядит следующим образом. В разделе 2 описывается базовая схема используемого частотно-временного артикуляторного синтезатора. Раздел 3 посвящён модели взаимодействия подсвязочной области и речевого тракта. В разделе 4 строится новая схема синтезатора, учитывающая влияние трахеи, бронхов и лёгких на характеристики тракта. Раздел 5 описывает эксперименты по синтезу некоторых звуко сочетаний с помощью построенной схемы. Наконец, раздел 6 содержит обсуждение полученных результатов.

2. Частотно-временной артикуляторный синтезатор речи

Частотно-временной синтезатор, используемый в данной статье, содержит несколько блоков (рис. 1).

Синтезатор работает на частоте дискретизации 16 кГц с квантованием каждого значения амплитуды на 16 бит. На вход синтезатора через одинаковые интервалы времени Δ (обычно полагаемые равными 10 мс) подаются значения площади поперечного сечения S и длины речевого тракта, а также значения некоторых параметров управления моделью площади голосовой щели $S_g(n)$ (n — номер дискретного отсчёта времени), включая контур основного тона. При этом значения площади тракта могут быть либо вычислены по некоторой математической модели артикуляции (например, с помощью алгоритма из [Баден и др., 2005]), либо получены путём непосредственного измерения с помощью ядерно-магнитной томографии [Story et al., 1998]. В качестве модели площади голосовой щели используется модель из [Сорокин, Макаров, 2008].

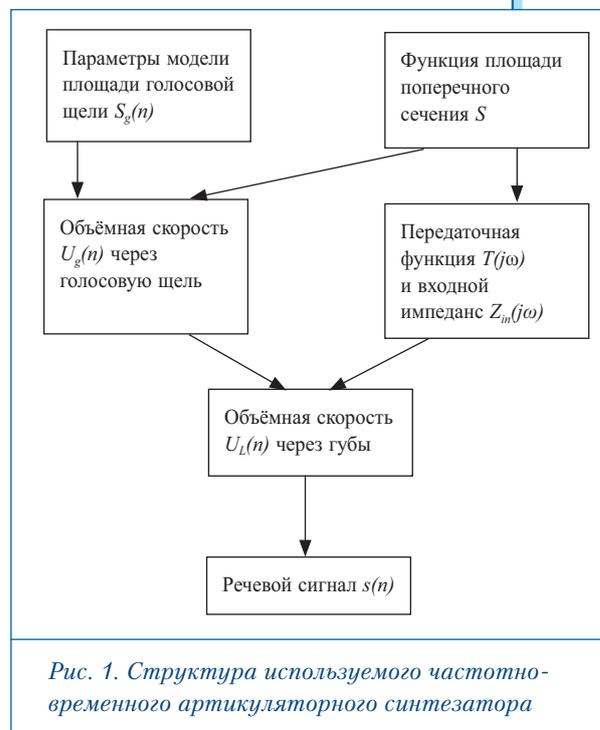


Рис. 1. Структура используемого частотно-временного артикуляторного синтезатора

Затем по функции площади речевого тракта определяется его передаточная функция $T(j\omega)$ и некоторые импедансы (в частности, входной акустический импеданс $Z_{in}(j\omega)$ в речевой тракт со стороны голосовой щели) с помощью модифицированного метода длинной линии [Макаров, 2009]. Кроме того, на этом этапе определяются координата и значение площади минимального сужения в тракте. Эти данные совместно с текущими значениями управляющих параметров площади $S_g(n)$ используются для вычисления объёмных скоростей и давления в различных участках речевого тракта, в т.ч. объёмной скорости $U_g(n)$, протекающей через голосовую щель, методом из [Hanson, Stevens, 2002].

На следующем этапе вычисляется объёмная скорость $U_L(n)$, протекающая через губы. Обычно для решения этой задачи используется один из двух методов. Первый метод заключается в вычислении по $T(j\omega)$ импульсной характеристики речевого тракта путём обратного преобразования Фурье с дальнейшей свёрткой этой характеристики с функцией $U_g(n)$ [Sondhi, Schroeter, 1987]. Во втором методе по $T(j\omega)$ сначала оцениваются её полюсы и вычеты для каждого полюса. Объёмная скорость у губ вычисляется средствами параллельного артикуляторно-формантного синтезатора [Lin, 1995]. На наш взгляд, эти методы не исчерпывают всех возможностей вычисления $U_L(n)$. Схема, описываемая в разделе 4, является ещё одним методом решения данной задачи.

Наконец, в последнем блоке вычисляется речевой сигнал $s(n)$ путём соответствующей фильтрации $U_L(n)$.

Как указано в разделе 1, описанная схема не позволяет учитывать влияние подсвяточной области на акустические характеристики речевого тракта. Соответствующая модификация схемы будет построена в разделе 4.

3. Взаимодействие подвязочной области и речевого тракта

Рассмотрим математическую модель, описывающую взаимодействие подвязочной области с передаточной функцией речевого тракта. Пусть T — передаточная функция речевого тракта, вычисленная в предположении бесконечного акустического импеданса голосовой щели, Z_{in} — входной акустический импеданс в речевой тракт со стороны голосовой щели, Z_{sub} — входной акустический импеданс в трахею со стороны голосовой щели, Z_g — акустический импеданс голосовой щели. Тогда передаточная функция T_{tr} речевого тракта с учётом конечного импеданса голосовой щели и наличия подвязочной области определяется так [Chi, Sonderegger, 2007]:

$$T_{tr} = T \frac{Z_g}{Z_{in} + Z_g + Z_{sub}}. \quad (1)$$

Функции T и Z_{in} определялись с помощью обобщённой схемы длинной линии по площадям поперечных сечений, измеренным с помощью магнитно-резонансной томографии речевого тракта реального диктора для английских гласных /A, E, I, U/ [Baer et al., 1991]. Во всех расчётах учитывалось наличие потерь на вязкое трение и теплопроводность, а также податливость стенок речевого тракта. Числовые значения всех необходимых параметров указаны в [Макаров, 2009].

В качестве модели подвязочной области использовалась полость объёма V_l , аппроксимирующая лёгкие, которая сочленялась с однородной трубой с потерями и податливыми стенками (трахея). Импеданс полости определялся так:

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

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

Акустический импеданс Z_g голосовой щели вычислялся по формуле, приведённой в [Сорокин, 1985]:

$$Z_g = \frac{12\mu h_g}{l_g d_g^3} + \frac{\rho_0 U_g}{S_g} + \frac{\rho_0 h_g}{S_g}. \quad (3)$$

Здесь μ — коэффициент вязкости воздуха, h_g — толщина голосовых складок, l_g — длина голосовых складок, d_g — ширина голосовой щели, U_g — среднее за период значение объёмной скорости, протекающей через голосовую щель, S_g — среднее за период значение площади голосовой щели. Предполагалось, что голосовая щель с достаточной степенью точности аппроксимируется прямоугольником, площадь которого может быть вычислена как $S_g = d_g l_g$.

Числовые значения параметров лёгких, трахеи и голосовой щели указаны в таблице 1.

Таблица 1

Параметры лёгких, трахеи и голосовых складок

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

Тестирование описанной модели взаимодействия подсвязочной области и речевого тракта проводилось для пяти значений площади голосовой щели S_g : 0 см² (случай бесконечно-го импеданса голосовой щели), 0.04 см², 0.08 см², 0.12 см², 0.2 см² (случай максимально открытой голосовой щели). На рис. 2 представлены амплитудно-частотные характеристики четырёх гласных звуков для указанных значений площади голосовой щели. Видно, что по мере раскрытия голосовой щели появляются дополнительные резонансы и антирезонансы, а основные резонансы сдвигаются по частоте.

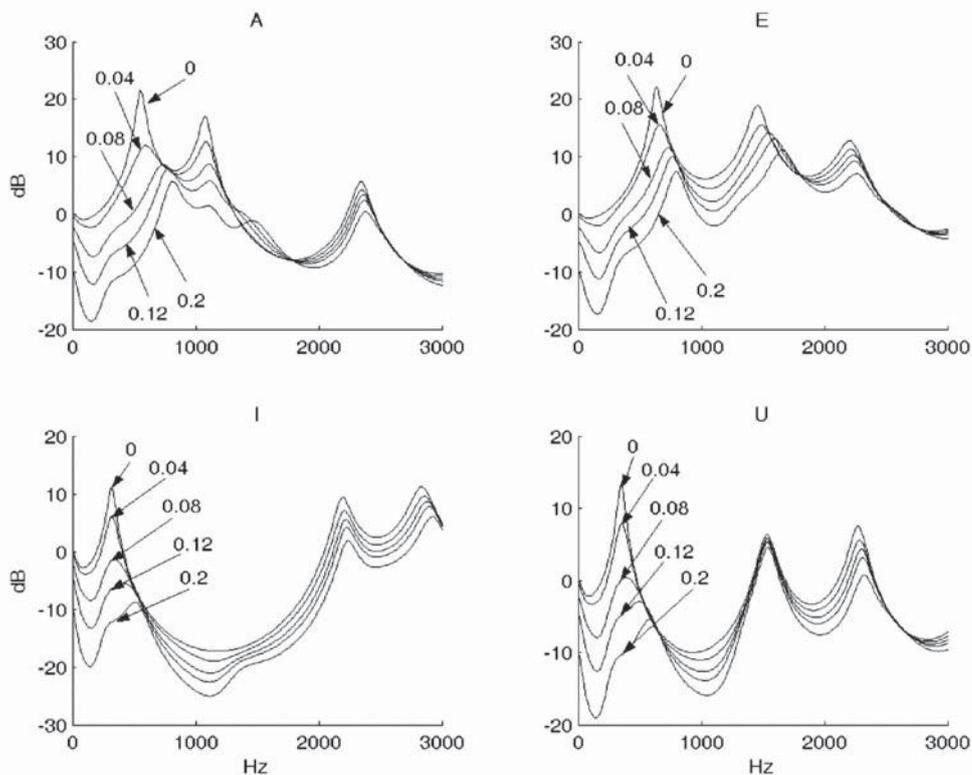


Рис. 2. Огибающие амплитудно-частотных характеристик для четырёх гласных звуков и пяти значений площади голосовой щели



Частоты основных резонансов определялись по пикам соответствующих амплитудно-частотных характеристик. Девиации F_1, F_2, F_3 (в %) по отношению к частотам $F_1^{(0)}, F_2^{(0)}, F_3^{(0)}$ для сомкнутых голосовых складок указаны в таблицах 2, 3 и 4 соответственно.

Таблица 2

Девиация первой формантной частоты

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

Таблица 3

Девиация второй формантной частоты

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

Таблица 4

Девиация третьей формантной частоты

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

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

4. Модель подсвяточной области в частотно-временном артикуляторном синтезаторе

Из соотношений (1)–(3) следует, что наиболее естественно влияние подсвяточной области на акустические характеристики речевого тракта моделируется в частотной области. Это соображение подсказывает следующую схему синтеза: сначала вычисляется последовательность кратковременных спектров Фурье от функции $U_g(n)$ в окне анализа $W(n)$ со сдвигом окна в R отсчётов. Затем передаточная функция речевого тракта пересчитывается согласно соотношению (1) и умножается на кратковременный спектр Фурье объёмной скорости $U_g(n)$. Синтез сигнала во временной области осуществляется методом наложения с суммированием. На последнем этапе сигнал пропускается через фильтр, моделирующий эффект излучения на губах.

Рассмотрим описанную схему более подробно. На вход алгоритма поступают: последовательность отсчётов объёмной скорости $U_g(n)$ и набор передаточных функций $T_k(j\omega)$ и акустических импедансов $Z_{in,k}(j\omega)$, вычисляемых через равные интервалы времени Δ по площади поперечного сечения речевого тракта.

Алгоритм выглядит следующим образом.

Задаём кратковременное окно анализа $W(n)$ и сдвиг окна, равный R отсчётам. Параметр R определяет, насколько часто будет вычисляться спектр Фурье от объёмной скорости, протекающей через голосовую щель.

Вычисляем последовательность дискретных кратковременных спектров Фурье от функции $U_g(n)$:

$$\tilde{U}_r(j\omega_p) = \sum_{n=0}^{N-1} U_g(n)W(rR - n)\exp(-j\omega_p n). \quad (4)$$

В этом соотношении N — длина окна анализа в отсчётах; ω_p — дискретная сетка отсчётов по частоте ($\omega_p = 2\pi p/N$, $0 \leq p \leq N-1$); $\{rR\}$ — дискретная временная сетка, на которой вычисляются функции $\tilde{U}_r(j\omega_p)$, $r = 0, 1, 2, \dots$

В каждом окне анализа вычисляем средние значения объёмной скорости через голосовую щель и площади голосовой щели. Соответствующие массивы обозначим как $u_g(r)$ и $s_g(r)$.

По $u_g(r)$, $s_g(r)$ вычисляем акустический импеданс голосовой щели $Z_{g,r}(j\omega_p)$ согласно (3). Отметим, что импеданс $Z_{g,r}(j\omega_p)$ может быть вычислен (для заданных параметров лёгких и трахеи) только один раз и в дальнейшем загружаться из файла, а не вычисляться каждый раз заново.

Интерполируем функции $T_k(j\omega)$ и $Z_{in,k}(j\omega)$, изначально заданные на временной сетке $\{k\Delta\}$, на новую временную сетку $\{rR\}$, $k, r = 0, 1, 2, \dots$

По функциям $T_r(j\omega)$ и $Z_{in,r}(j\omega)$, $Z_{g,r}(j\omega_p)$ и $Z_{g,r}(j\omega_p)$ вычисляем передаточную функцию $\tilde{T}_r(j\omega_p)$, учитывающую влияние подсвяточной области, по формуле (1).

Вычисляем речевой сигнал методом наложения с суммированием:

$$s(n) = \sum_{r=0}^{+\infty} \left[\frac{1}{N} \sum_{p=0}^{N-1} I(j\omega_p) \tilde{T}_r(j\omega_p) \tilde{U}_r(j\omega_p) \exp(j\omega_p n) \right] \quad (5)$$

В данном соотношении множитель I моделирует импеданс излучения на губах.

Выбор окна анализа $W(n)$ определяется двумя требованиями. С одной стороны, длительность L_t окна должна быть как можно короче для достижения хорошего разрешения по времени. С другой стороны, ширина L_w главного лепестка спектра Фурье окна $W(n)$ также должна быть как можно уже для достижения хорошего разрешения по частоте. Хорошо известно, что эти требования противоречат друг другу в силу принципа неопределённости Гейзенберга:

$$L_t L_w \geq \frac{1}{2}. \quad (6)$$

Точную нижнюю грань неравенству (6) доставляет временное окно Гаусса [Штарк, 2007]. Оно определяется следующей формулой:

$$W(n+1) = \exp\left(-\frac{1}{2} \left[2.5 \frac{n - N/2}{N/2}\right]^2\right). \quad (7)$$

Иными словами, окно Гаусса обеспечивает наилучший компромисс между разрешением по времени и частоте. В силу этого обстоятельства, оно и было выбрано в качестве временного окна анализа. Длительность окна полагалась равной 10 мс (160 отсчётов при частоте дискретизации 16 кГц). Для дополнительного повышения разрешения по частоте сигнал $U_g(n)$ в окне анализа дополнялся нулями до 512 отсчётов.

Выбор параметра R определяется шириной главного лепестка L_w спектра Фурье от функции $W(n)$. Пусть $L_w = 2\pi f_c$, где f_c — значение граничной частоты главного лепестка в Гц. Тогда параметр R должен удовлетворять следующему неравенству [Рабинер, Шафер, 1981]:

$$R \leq 1 / (2f_c). \quad (8)$$

Несоблюдение этого условия может привести к существенным искажениям сигнала во временной области при вычислении обратного преобразования Фурье по формуле (5). В проведённом исследовании параметр R полагался равным 2 мс (32 отсчёта при частоте дискретизации 16 кГц).

Формула (5) является классическим соотношением для синтеза сигнала по последовательности его кратковременных спектров Фурье методом наложения с суммированием [Рабинер, Шафер, 1981]. Известны различные модификации этой формулы [Griffin, Lim, 1984; Veldhuis, He, 1996], основанные на введении кратковременных окон синтеза и различных нормировок сигнала. Мы реализовали все модификации, однако не обнаружили никакой перцептивной разницы между сигналами, синтезированными по разным схемам. Поэтому в дальнейшем мы обсуждаем только результаты, полученные на основе формулы (5).

5. Эксперименты

Данными для экспериментов послужили площади поперечного сечения, основанные на результатах ядерно-магнитной томографии речевого тракта реального диктора [Story, 2005] и включавшие в себя произношения отдельных гласных и различных звукосочетаний. В дальнейшем анализе использовались последовательности площадей сечения для гласных /A, U, I/, а также для звукосочетаний /AU/ и /IO/. Исходные акустические файлы для данных звуков были недоступны, так что мы не могли сопоставить синтезированные сигналы с их исходным звучанием.

Сначала была проверена гипотеза о перцептивной эквивалентности стандартных схем частотно-временного синтезатора и схемы из раздела 4 без учёта влияния подсвязочной области. Для этого сигналы были синтезированы с помощью схем из [Sondhi, Schroeter, 1987; Lin, 1995], а также схемы, построенной выше. Чтобы исключить влияние подсвязочной области, вместо функции $\tilde{T}_r(j\omega_p)$ в формуле (5) использовалась функция $T_r(j\omega)$. На слух сигналы оказались неотличимыми друг от друга.

Затем были синтезированы речевые сигналы с учётом подсвязочной области, параметры для которой заимствовались из таблицы 1. Прослушивание результирующих сигналов на бюджетной бытовой аудиоаппаратуре не выявило между ними существенной разницы. При прослушивании этих же звуков на высококачественной студийной аппаратуре (мониторах KRK VXT4 или головных телефонах AKG K240 mk2, подключённых к аудиоинтерфейсу E-MU 0404 USB, и другой аппаратуре более высокого класса) разница между ними оказалась весьма заметной. При этом сигналы, синтезированные с учётом влияния трахеи, бронхов и лёгких, имели более сочный и бархатистый оттенок по сравнению с сигналами, синтезированными без подсвязочной области.

Напомним, что в данной работе не ставилась задача выяснения роли подсвязочной области для восприятия речи. Решение такой задачи требует особого исследования.

6. Обсуждение

Цель проведённого исследования заключалась в создании схемы артикуляторного синтезатора, моделирующей взаимодействие речевого тракта и подсвязочной области. Построенная схема обладает двумя существенными преимуществами перед стандартными алгоритмами артикуляторного синтеза. Первое заключается в том, что влияние трахеи, бронхов и лёгких на передаточную функцию речевого тракта моделируется в частотной области. Это позволяет учесть нелинейные функции частоты в переносном импедансе трахеи наиболее адекватным образом, без упрощающих предположений, характерных для временных синтезаторов. Вторым преимуществом является использование кратковременного анализа Фурье и метода наложения с суммированием для синтеза сигнала. Выбирая достаточно высокую частоту вычисления спектра Фурье (или, что эквивалентно, увеличивая значение параметра R), можно синтезировать быстрые изменения спектральных характеристик речевого сигнала (что практически нереализуемо в стандартных методах частотно-временного артикуляторного синтеза).

В настоящем исследовании соотношение (5) реализовано посредством вычисления обратного быстрого преобразования Фурье от спектров звуков. Возможны иные реализации этого соотношения. Например, раскладывая комплексные экспоненты в (5) по синусам и косинусам, можно осуществить синтез средствами гармонического вокодера



[McAulay, Quatieri, 1986; George, Smith, 1997]. Соотношение (5) может быть реализовано методами, разработанными в рамках модуляционной модели речевого сигнала [Potamianos, Maragos, 1999]. Преимущество подобных схем перед схемой, описанной в разделе 4, очевидно в методе синтеза речевой волны синхронно с основным тоном. В этом случае длина окна анализа N становится равной длительности текущего периода основного тона в отсчётах, что приводит к значительным ускорениям в вычислениях и уменьшению объёмов памяти. Например, для частоты основного тона 100 Гц и частоты дискретизации 16 кГц требуется вычислять передаточную функцию тракта и акустические импедансы лишь для 80 гармоник (до частоты Найквиста), в то время как метод, построенный в разделе 4, требует вычисления акустических характеристик тракта для 512 спектральных отсчётов.

Несмотря на то, что схема была построена для учёта взаимосвязи речевого тракта и подсвяточных областей, она может эффективно использоваться для синтеза всех явлений акустики речеобразования, наиболее адекватно моделируемых именно в частотной области. В качестве примера можно упомянуть турбулентный шум, служащий основным источником звука для фрикативных согласных. Известно [Zhang et al., 2002], что многие характеристики турбулентного шума существенно зависят от трёхмерной геометрии шумящего отверстия, наличия/отсутствия препятствия на пути воздушного потока, а также от угла падения потока на рассеивающее препятствие. Практика показывает, что адекватно моделировать эти явления значительно проще в частотной, а не во временной области [Narayanan, Alwan, 2000]. В качестве другого примера можно указать «обратное» влияние входного импеданса в речевой тракт со стороны голосовой щели на воздушный поток, протекающий через гортань. Во временной области это влияние моделируется свёрткой данного импеданса с отсчётами объёмной скорости $U_g(n)$. Технические трудности данного подхода подробно обсуждаются в [Sondhi, Schroeter, 1987]. С другой стороны, в частотной области это влияние сводится к произведению входного акустического импеданса и спектра Фурье от объёмной скорости через голосовую щель.

Литература

1. Сорокин В.Н. Теория речеобразования. М.: Радио и Связь, 1985. 312 с.
2. Stevens K. Acoustic Phonetics. The MIT Press, 2000. 614 p.
3. Chi X., Sonderegger M. Subglottal Coupling and its Influence on Vowel Formants // J. Acoust. Soc. Amer. 2007. Vol. 122, P. 1735–1745.
4. Сорокин В.Н. Синтез речи. М.: Наука, 1992. 392 с.
5. Assmann P., Katz W. Synthesis Fidelity and Time-Varying Spectral Change in Vowels // J. Acoust. Soc. Amer. 2005. Vol.117. P. 886–895.
6. Морозов В.П. Искусство резонансного пения. М., 2002. 496 с.
7. Kawahara H., Masuda-Katsue I. and Cheveigne A.de. Restructuring Speech Representations Using a Pitch-Adaptive Time-Frequency Smoothing and an Instantaneous — Frequency Based F0 Extraction: Possible Role of a repetitive Structure in Sounds // Speech Communication. 1999. Vol.27. P. 187–207.

8. Баден П., Макаров И.С., Сорокин В.Н. Алгоритм вычисления площадей поперечного сечения речевого тракта // Акуст. журнал. 2005. Т. 51. No. 1. С. 52–58.
9. Story B., Titze I. and Hoffman E. Vocal Tract Area Functions for an Adult Female Speaker Based on Volumetric Imaging // J. Acoust. Soc. Amer. 1998. Vol.104. P. 471–487.
10. Сорокин В.Н., Макаров И.С. Определение пола диктора по голосу // Акуст. журнал. 2008. Т. 54. No. 4. С. 659–668.
11. Макаров И.С. Аппроксимация речевого тракта коническими рупорами // Акуст. журнал. 2009. Т. 55. No. 2. С. 256–265.
12. Hanson H., Stevens K. A Quasarticulatory Approach to Controlling Acoustic Source Parameters in a Klatt-Type Formant Synthesizer Using HLsyn // J. Acoust. Soc. Amer. 2002. Vol. 112. P. 1158–1182.
13. Sondhi M.M., Schroeter J. A Hybrid Time-Frequency Domain Articulatory Synthesizer // IEEE Trans. Acoust., Speech, Signal Process. 1987. Vol. ASSP-35. No. 7. P. 955–967.
14. Lin Q. A Fast Algorithm for Computing the Vocal Tract Impulse Response from the Transfer Function // IEEE Trans. Speech, Audio Process. 1995. Vol. 3. No. 6. P. 449–457.
15. Штарк Г. Применение вейвлетов для цифровой обработки сигналов. М.: Техносфера, 2007. 192 с.
16. Рабинер Л., Шафер Р. Цифровая обработка речевых сигналов. М.: Радио и Связь, 1981. 496 с.
17. Griffin D., Lim J. Signal Estimation From Modified Short-Time Fourier Transform // IEEE Trans. Acoust., Speech, Signal Process. 1984. Vol. 32. No. 2. P. 236–243.
18. Veldhuis R., He H. Time-Scale Pitch Modification of Speech Signals and Resynthesis From the Discrete Short-Time Fourier Transform // Speech Communication. 1996. Vol. 18. P. 257–279.
19. Story B. A Parametric Model of the Vocal Tract Area Function for Vowel and Consonant Simulation // J. Acoust. Soc. Amer. 2005. Vol. 117. P. 3231–3254.
20. McAulay K., Quatieri T. Speech Analysis / Synthesis Based On a Sinusoidal Representation of Speech // IEEE Trans. Acoust., Speech, Signal Process. 1986. Vol. 34. P. 744–754.
21. George E., Smith M. Speech Analysis / Synthesis and Modification Using an Analysis — by — Synthesis / Overlap — Add Sinusoidal Model // IEEE Trans. Speech, Audio Process. 1997. Vol.5. No.6. P. 389–406.
22. Potamianos A., Maragos P. Speech Analysis and Synthesis Using an AM-FM Modulation Model // Speech Communication. 1999. Vol.28. P. 195–209.
23. Narayanan S., Alwan A. Noise Source Models for Fricative Consonants // IEEE Trans. Speech, Audio Process. 2000. Vol.8. No.2. P. 328–344.
24. Zhang Zh., Mongeau L. and Frankel S. Broadband Sound Generation by Confined Turbulent Jets // J. Acoust. Soc. Amer. 2002. Vol.112. P. 677–689.

Горбунов К.С.,

Макаров И.С.,

*кандидат технических наук, старший научный сотрудник
Института Проблем Передачи Информации им. А.А. Харкевича РАН.
speechprod_mak@mail.ru*