

УДК 612.789+616.89-008.4341-7

НЕЛИНЕЙНАЯ МОДЕЛЬ АКУСТИЧЕСКОГО ГОЛОСОВОГО ИСТОЧНИКА

© 1997 г. О. П. Спяров

Санкт-Петербургский НИИ уха, горла, носа и речи

198013 Санкт-Петербург, ул. Бронницкая, 9

Поступила в редакцию 11.07.95 г.

Предлагается модель голосового источника, учитывающая нелинейные эффекты голосообразования путем суммирования всего эквизистантного спектрального ряда. Показана связь речевого удара с особенностями динамики воздушного потока через голосовые складки. Установлено, что в первом порядке гидродинамической теории возмущений имеет место бифуркация Хопфа. Демонстрируется соответствие предлагаемой модели с экспериментальными данными.

Согласно классической теории речеобразования Фанта [1, 2], объемная скорость $u(t)$ воздушного потока на выходе голосового источника (щели, образованной колеблющимися голосовыми связками в латеральной (поперек горла) плоскости) представляет собой периодическую (с периодом основного тона $T_0 = 1/F_0$, где F_0 – частота основного тона, изменяющаяся в диапазоне 100–200 Гц) последовательность импульсов. Особенностью каждого импульса в этой последовательности есть скошенность заднего фронта импульса, т.е. более быстрое, чем синусоидальное на частоте F_0 , его спадание. Кроме того, каждый импульс характеризуется амплитудой u_0 – скоростью потока при максимально открытой щели, зависящей от перепада давления ΔP на входе и выходе из щели. Передний фронт предполагается нарастающим достаточно плавно.

С такой картиной воздушного потока в теории Фанта соотносится соответствующая спектральная картина звукового давления на выходе голосовой щели. Этот спектр является одной из основных характеристик при описании акустического речевого сигнала. Однако теоретическое описание спектра сталкивается с определенными трудностями. Дело в том, что в линейной (по скорости) гидродинамической теории звука ($u/c \ll 1$, где c – скорость звука) вводится понятие акустического потенциала ϕ , удовлетворяющего однородному (без правой части) волновому уравнению, т.е. ϕ представляет собой потенциал в поле без объемных источников: вся энергия попадает в поле благодаря условию непрерывности нормальной составляющей скорости на поверхности колеблющихся в этом поле тел. Если же не пренебрегать нелинейными поправками по скорости u , то в правой части волнового уравнения возникает ряд теории возмущений по степеням параметра малости u/c , и эта правая часть (в виде ряда)

интерпретируется как появление объемных источников. Функция ϕ остается потенциалом, т.е. не теряет своих потенциальных свойств, но это уже потенциал в поле с объемными источниками.

В силу своих потенциальных свойств потенциал $\phi(\vec{r}, t)$ в пространственной точке \vec{r} связан с вектором скорости $\vec{u}(\vec{r}, t)$ и звуковым давлением $p(\vec{r}, t)$ обычными соотношениями $\vec{u}(\vec{r}, t) = \text{grad}\phi(\vec{r}, t)$ и $p(\vec{r}, t) = \dot{\phi}(\vec{r}, t)$, где точка означает производную по времени, так что, при условии разделения временных и пространственных координат, для зависящих только от времени компонент $p(t)$ и $u(t)$ получаем соотношение:

$$p(t) \sim \dot{u}(t), \quad (1)$$

но в поле с наличием объемных источников.

Следовательно, в силу импульсной периодичности функции $\dot{u}(t)$ функция $u(t)$ (или, что то же самое $p(t)$) будет носить также импульсный, периодический (с периодом T_0) характер.

Далее, согласно теории Фанта, в силу указанной периодичности спектр звукового давления $p(t)$ представляется в виде эквидистантного набора гармоник на частотах nF_0 с некоторым образом заданным законом убывания их амплитуд $q(n)$, где n – номер гармоники. Из общей теории разложения в ряд Фурье ясно, что спектральный наклон $q(n)$ связан со спектральным составом одиночного импульса $p(t)$ или $\dot{u}(t)$, и в силу эффекта скошенности (skewness), главным образом будет определяться крутизной заднего фронта (величиной отрицательной производной) на участке заднего фронта импульса $u(t)$. Временную зависимость участка заднего фронта $u(t)$ обозначим как $u_s(t)$, а область определения ее, приведен-

ную к первому периоду, зададим как $[T_0/4 - \Delta, T_0/2 - \Delta]$, где Δ – некая малая величина.

Оценить форму импульса $u(t)$, а следовательно, и спектральный закон $q(n)$ – весьма трудная задача, поддающаяся только компьютерному моделированию [3], так как приходится совместно решать нелинейную задачу теории упругости (с параметризацией коэффициента натяжения связок $\sigma(t)$) для определения площади щели $S(t)$ между колеблющимися складками, находящимися, в свою очередь, под воздействием аэродинамических сил в потоке со скоростью $u(t)$ и с перепадом связочного давления $\Delta P(t)$ (который также задается как параметр). Задача замыкается сложной гидродинамической зависимостью скорости $u(t)$ от площади $S(t)$ (в виде обобщенного уравнения Рикатти).

Нужно иметь в виду, что помимо сложности этой “пассивной” совместной задачи теории упругости и гидродинамики, задача становится необозримо сложной с учетом ее “активного” характера. Действительно, и коэффициент натяжения связок $\sigma(t)$ является функцией активности группы мышц, управляющих движением связок (так называемой синергии связок) [4], и $\Delta P(t)$ является функцией активности какой-то (пока неясно, даже какой) синергии мышц дыхания. Динамика совместного функционирования таких синергий, скорее работает на принципах самоорганизации сложных систем [5, 6], но детальная проработка проблемы еще достаточно далека от завершения.

Скорее всего, решение замкнутой задачи по отысканию спектральной зависимости $q(n)$ на сегодняшний день вряд ли возможно, поэтому на первый план выдвигаются методы моделирования.

Мы предлагаем промоделировать скошенность импульса скорости потока $u_s(t)$ на приведенном к первому периоду основного тона интервале $[T_0/4 - \Delta, T_0/2 - \Delta]$ ее спектром $q(n)$ в виде убывающей геометрической прогрессии q^n , где $0 \leq q < 1$. Тогда функция скорости потока $u_s(t)$ (приведенная к амплитуде потока, т.е. $\tilde{u}_s = (u_s - u_0)/u_0$) выразится в виде:

$$\tilde{u}_s(t) = \operatorname{Re}\{1 + qe^{2\pi i F_0 t} + q^2 e^{4\pi i F_0 t} + \dots\} = \operatorname{Re}[1 - e^{2\pi i F_0 t}]^{-1} \quad \text{при } T_0/4 - \Delta \leq t \leq T_0/2 - \Delta, \quad (2)$$

где q имеет физический смысл относительной (к S_{\max}) амплитуды гармонической (на частоте F_0) зависимости площади окна щели $S(t)$. Действительно, согласно Сорокину [3], исходящему из своих компьютерных расчетов, если представить $S(t) = q \sin(2\pi F_0 t)$, (где $q = S_0/S_{\max}$), то при $q \leq 0.1$ скорость $u(t)$ изменяется гармонически, что согласуется с разложением (2), затем при $q > 0.1$ влияние изменения площади окна $S(t)$ начинает

возрастать и функция скорости потока становится несимметричной.

Из разложения (2) мы видим, что с увеличением q увеличивается вклад высших гармоник в функцию скошенности $\tilde{u}_s(t)$. Убедимся, что этот вклад приводит к усилению скошенности, т.е. к увеличению (по абсолютной величине) отрицательной производной $\dot{u}_s(t)$. В качестве меры скошенности \hat{S} , с целью получения аналитического результата, возьмем производную в точке $t = T_0/4$ и вернемся к ненормированной амплитуде, тогда получаем

$$\hat{S} = -\dot{u}_s(T_0/4) = 2\pi F_0 u_0 f(q), \quad (3)$$

где функцию $f(q)$ представима в виде:

$$f(q) = \frac{2q^2}{(1+q^2)^2}. \quad (3')$$

Действительно, мы имеем выпуклую кривую, плавно нарастающую от 0 и асимптотически стремящуюся к пределу $f(q \rightarrow 1) = 0.5$.

Таким образом, поставленной цели – промоделировать скошенность заднего фронта импульса скорости, увеличивающуюся с нарастанием площади раскрытия окна q , разложение (2) удовлетворяет.

Следует сказать, что скошенность заднего фронта импульса скорости часто связывают с речевым ударением [7]. Таким образом, мы можем констатировать, что в нашей модели ударение \hat{S} определяется формулой (3). Если считать частоту основного тона $F_0(\tau)$, амплитуду скорости потока $u_0(\tau)$ относительную площадь окна $q(\tau)$ функциями “медленного” времени τ (сравнимым с несколькими периодами основного тона T_0), то для количественного выраженного ударения в таком времени мы получаем выражение:

$$\hat{S}(\tau) = 2\pi F_0(\tau) u_0(\tau). \quad (4)$$

Как мы уже говорили, частота основного тона есть функция времени $F_0(\tau)$, и, кроме установившегося режима, спектральная картина сигнала может проявлять различные переходные режимы. Если, начиная с какого-то момента, $F_0(\tau)$ начнет спадать до нуля, т.е. $T_0(\tau) \rightarrow \infty$, то щель раскроется и колебаний нет – это “мягкое” выключение. “Мягкое” включение идет в обратном порядке, при нарастании $F_0(\tau)$ от 0 до конечных значений. Возможны и “резкие” режимы: поскольку u_0 зависит от перепада давления ΔP , то резкое исчезновение ΔP может привести, в соответствии с (1) и (3), к исчезновению акустического поля, при этом скошенность нулевая – щель раскрыта. Эта же модель имитирует шум (при $T_0 \rightarrow 0, F_0 \rightarrow \infty$ имеет место последователь-

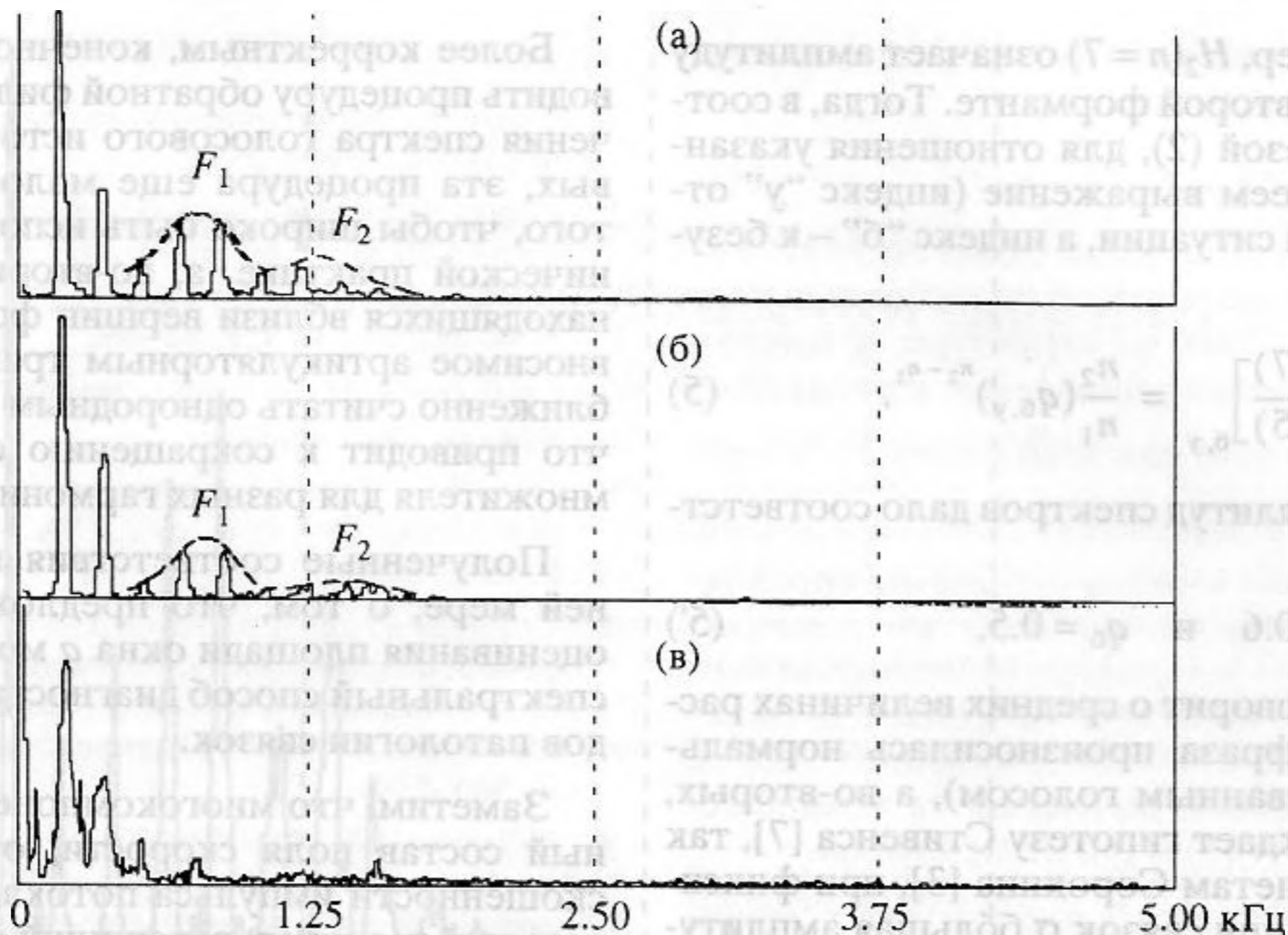


Рис. 1. Спектральные срезы для ударного “а” в ситуации “лала” в слове “запылала” в норме (а); безударного “а” в норме (б); ударного “а” в случае паралича гортани (в).

ность очень близких коротких импульсов потока через щель). Все эти режимы наблюдаются на эксперименте [6].

Далее, в согласии с теорией Фанта, полученное излучение с эквидистантным дискретным спектральным составом подвергается линейному пассивному фильтрующему воздействию последующего за связками речевого тракта. Эта фильтрация происходит на частотах так называемых формант F_1, F_2, F_3 с довольно низкими добротностями, в результате чего и получается спектр звукового давления, регистрируемого микрофоном. Этот спектр представляется в виде набора формант с амплитудами, убывающими по закону q^n , где n – номер ближайшей к вершине форманты гармоника голосового источника, и с крыльями, заполненными убывающими по амплитуде гармониками в соответствии с добротностями формант. Частоты формант F_1, F_2, F_3 в соответствии с оценками Фланагана [8], располагаются соответственно на частотах 500, 1500, 2500 Гц и далее с интервалом $c/2l$ Гц, $l = 17$ см (для мужского голосового тракта). Более тщательный учет резонансных свойств голосового тракта свидетельствует о раздвоении первой форманты по мере увеличения площади щели, так что возникающая дополнительная форманта $F_1^{(1)}$ пропускает первую гармонику голосового источника [3].

Для того чтобы проверить, согласуется ли гипотеза (2) с экспериментальными данными, нами были отобраны две спектральные картины ударного и безударного “а” в сочетании “лала” в слове “за-

пылала” из фразы стандартного текста “Заря запылала пожаром”. Спектрограммы измерялись с целью проверки гипотезы Стивенса и др. [7] о том, что усиление амплитуд H_1, H_2 высших формант F_1, F_2 может быть обусловлено не повышением частоты основного тона, а только лишь изменением разности давлений ΔP на связках. Полученная нами формула для ударения (4) допускала такую возможность, так как, в соответствии с нею, ударение изменяется не только за счет F_0 , но и амплитуды потока u_0 , в основном зависящей от периода давления ΔP . Отобранные два спектральных среза подтверждали гипотезу Стивенса. Из сравнения спектрограмм видно (см. рис. 1), что в ударной ситуации (рис. 1а) происходит усиление формант F_1 и F_2 по сравнению с безударной ситуацией (рис. 1б) без изменения частоты основного тона (расстояния между гармониками остаются без изменения).

В ходе обработки приведенных экспериментальных данных удалось оценить степень раскрытия голосовой щели q . Действительно, гипотеза (2) при учете соотношения (1) позволяет, производя сравнение амплитуд гармоник $p(t)$, определять величину q . Эта величина имеет диагностическое значение, например, при различных поражениях складок (различного рода параличи, операционные воздействия и т.д.). Счет гармоник ведется, начиная с нулевой в начале координат, слева направо ($n = 0, 1, 2, \dots$).

Для ударной и безударной ситуации было взято отношение амплитуд $H_2(n = 7)$ и $H_1(n = 5)$. Обо-

значение, например, $H_2(n=7)$ означает амплитуду 7-й гармоники во второй форманте. Тогда, в соответствии с гипотезой (2), для отношения указанных амплитуд имеем выражение (индекс "у" относится к ударной ситуации, а индекс "б" – к безударной):

$$\left[\frac{H_2(n_2=7)}{H_1(n_1=5)} \right]_{б,у} = \frac{n_2}{n_1} (q_{б,у})^{n_2-n_1}. \quad (5)$$

Измерение амплитуд спектров дало соответственно:

$$q_y = 0.6 \quad \text{и} \quad q_b = 0.5, \quad (5')$$

что, во-первых, говорит о средних величинах раскрытия связок (фраза произносилась нормальным, не форсированным голосом), а во-вторых, еще раз подтверждает гипотезу Стивенса [7], так как согласно расчетам Сорокина [3], при фиксированном натяжении связок σ большая амплитуда раскрытия щели соответствует большему перепаду под- и надсвязочного давления ΔP . То, что фиксированное значение F_0 (так в нашем эксперименте) соответствует фиксированному натяжению связок σ , подтверждается теоретически Стори и Титцем [4] и экспериментально Бэкман и др. [9], которая установила строгую корреляцию между F_0 и уровнем сокращения перстнечитовидной мышцы, определяющим натяжение связок σ .

То, что полученные значения q не являются случайными, а отражают закономерность в формировании спектрального закона, было подтверждено также тем, что в соответствии с гипотезой (2) должно выполняться соотношение:

$$\frac{\left[\frac{H_1(n=1)}{H_2(n=4)} \right]_y}{\left[\frac{H_1(n=1)}{H_2(n=4)} \right]_б} = \left(\frac{q_b}{q_y} \right)^3. \quad (6)$$

Левая часть соотношения (6) определялась непосредственно из спектров и дала значение 0.6, а правая – из вычисленных ранее значений q_y и q_b (см. равенства (5')) и дала соответственно 0.57.

Нами был проведен анализ спектра второго безударного "а" в слове "запылала", произнесенного пациенткой с диагнозом: "двусторонний паралич гортани". Оценка степени раскрытия окна q , проведенная по формуле (1) для первой и четвертой гармоники первой форманты (см. рис. 1в) дает величину $q \cong 0.3$, что значительно меньше соответствующего параметра для такой же ситуации в норме ($q \cong 0.5$). Такое значительное уменьшение площади раскрытия окна q может свидетельствовать о патологических изменениях в голосовом аппарате, происшедших вследствие паралича.

Более корректным, конечно, было бы производить процедуру обратной фильтрации для получения спектра голосового источника, но, во-первых, эта процедура еще мало разработана для того, чтобы широко быть использованной в клинической практике, а, во-вторых, для гармоник, находящихся вблизи вершин формант, затухание, вносимое артикуляторным трактом, можно приближенно считать однородным по формантам [8], что приводит к сокращению соответствующего множителя для разных гармоник в формуле (5).

Полученные соответствия говорят, по крайней мере, о том, что предложенная процедура оценивания площади окна q может дать удобный спектральный способ диагностики некоторых видов патологии связок.

Заметим, что многокомпонентный спектральный состав поля скорости, отражающий факт скошенности импульса потока, носит фундаментальный характер, говорящий о сугубо нелинейности соответствующей аэродинамической (или гидродинамической) задачи. Эта нелинейность требует от исследователя и новых, адекватных методов исследования спектрального состава излучения. Попытаемся это продемонстрировать на конкретных примерах. То, что ряд (2) представляет решение нелинейной гидродинамики, легко убедиться, дважды продифференцировав (2) по времени. В результате получаем:

$$\frac{\partial^2 \tilde{u}_s(t)}{\partial t^2} + (2\pi F_0)^2 u_s(t) = F(u_s^2). \quad (7)$$

Структура волнового уравнения для плоской волны с источниками $F(u_s^2)$ (где F – некоторая функция) в правой части уравнения (7) – сугубо нелинейная ситуация.

Если бы мы оставили в разложении (2) только два первых слагаемых (гармонический процесс на частоте F_0), а остальную часть ряда рассматривали бы как малое возмущение, то получили бы стандартную ситуацию, исследованную в гидроакустике, а именно для малой поправки u_1 в представлении скорости $u = u_0 + u_1$, где u_0 – колебание на частоте основного тона, в силу нелинейности уравнений гидродинамики получили бы уравнение [10]:

$$\dot{u}_1 = \alpha u_1 (1 - \beta |u_1|^2), \quad (8)$$

где α и β – комплексные коэффициенты, зависящие от числа Рейнольдса R , определяющего вязкость среды. Такое уравнение имеет решение, обладающее бифуркациями Хопфа [10, 11]. Переход к хаосу для такого уравнения, обычно связанный с ростом вязкости, происходит по сценарию Рюэля–Тakens, который приводит к воз-

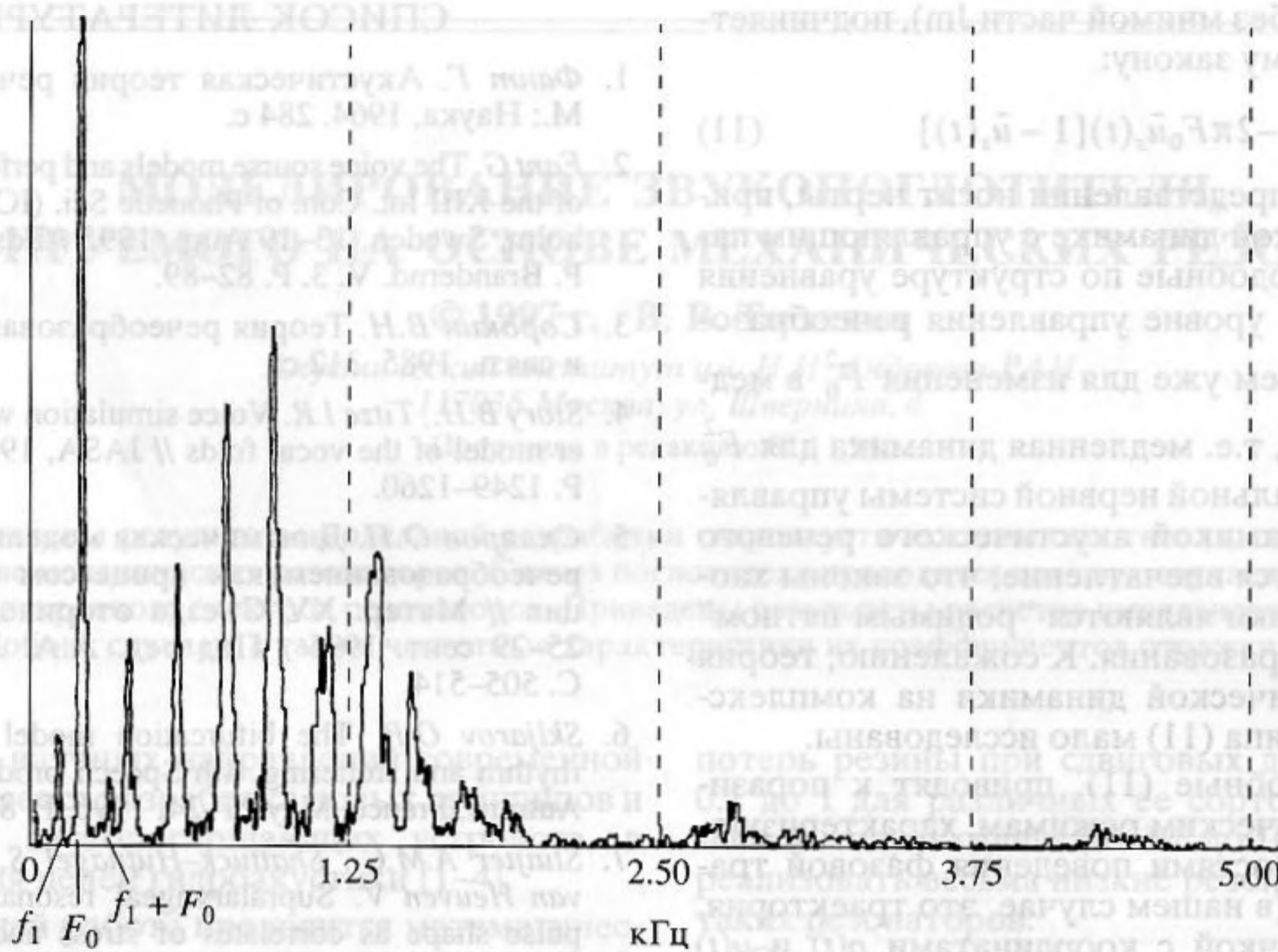


Рис. 2. Добавочные сателлиты на частотах f_1 и $f_1 + F_0$ в эквидистантном спектре голосового источника.

никновению дополнительной частоты f_1 , связанной с частотой основного тона F_0 соотношением:

$$\frac{f_1}{F_0} = Q, \tag{9}$$

где Q – иррациональное число, равное так называемому “золотому среднему” $Q = (\sqrt{5} - 1)/2 = 0.618339$ [11]. Происходит бесконечная “намотка на тор” – пример странного аттрактора. Интересно, что эту частоту, составляющую ≈ 0.6 от F_0 , действительно, можно наблюдать в спектрах (см. рис. 2). На рис. 2 между нулевой и первой гармониками видим колебание с малой амплитудой, хорошо согласующиеся с формулой (9). Появление подобных спектральных составляющих, вероятно, происходит на этапе раскрытия (разлипания) окна, когда во-первых, $q \ll 1$ (так как основную роль играет лишь первая гармоника F_0), а, во-вторых, роль гидродинамических эффектов в слизистом слое, покрывающем поверхность связок [4], с большими, чем в воздухе, свойствами вязкости. Число Рейнольдса R растет, и, в соответствии со сценарием Рюэля–Такенса для уравнения (8) возможно появление гармоники f_1 в отношении “золотого среднего” к основной гармонике F_0 .

Видная на рис. 2 добавочная компонента между первой и второй гармониками также находит объяснение в рамках теории возмущений. Дело в том, что добавочное поле за счет объемных ис-

точников можно представить в виде интеграла от произведения скоростей $u = u_0 + u_1$ по пространственной области, где поправки существенны [10]:

$$p(\vec{r}_1, t) \frac{\rho}{4\pi} \int \frac{\partial^2 T_{ik}(\vec{r}_1, t)}{\partial x_{1i} \partial x_{1k}} \Big|_{r=\frac{R}{c}} \frac{(dv_1)}{R}, \tag{10}$$

здесь \vec{r} – вектор точки наблюдения; \vec{r}' – радиус-вектор точки интегрирования; $R = |\vec{r} - \vec{r}'|$; ρ – плотность воздуха; T_{ik} – так называемый тензор скоростей $T_{ik} = u_i u_k$, где u_i и u_k – компоненты вектора скорости \vec{u} . Такое представление допускает возникновение колебаний на комбинационной частоте $Q_0 + f_1$, которую мы, по всей видимости, и наблюдаем на эксперименте.

Показанный пример иллюстрирует, насколько скуден был бы спектр голосового источника, если бы он описывался теорией возмущений в первом порядке по нелинейности (первая гармоника и два незначительных сателлита). Поэтому мы подчеркиваем, что гипотеза (2) позволяет просуммировать весь ряд теории возмущений (причем при параметрах $q \sim 1$) – ситуация, крайне редкая для сильных ($q \sim 1$) нелинейных взаимодействий, и нелинейных задач акустики – в частности.

Интересно отметить, что динамика комплексной величины $\tilde{u}_s(t)$, как легко видеть, продиффе-

ренцировав (2) (без мнимой части Jm), подчиняется логистическому закону:

$$\dot{\bar{u}}_s(t) = -2\pi F_0 \bar{u}_s(t) [1 - \bar{u}_s(t)] \quad (11)$$

и в дискретном представлении носит черты, присущие хаотической динамике с управляющим параметром F_0 . Подобные по структуре уравнения мы получаем на уровне управления речеобразованием [6], причем уже для изменения F_0^2 в медленном времени, т.е. медленная динамика для F_0^2 на уровне центральной нервной системы управляет быстрой динамикой акустического речевого сигнала. Создается впечатление, что законы хаотической динамики являются "родимым пятном" процесса речеобразования. К сожалению, теория уравнений хаотической динамики на комплексной плоскости типа (11) мало исследованы.

Законы, подобные (11), приводят к поразительным динамическим режимам, характеризующимся особенностями поведения фазовой траектории задачи (в нашем случае, это траектория, описываемая точкой с координатами $p(t)$ и $u(t)$ при изменении управляющего параметра (в данном случае — F_0). Особенности эти заключаются в наличии особых зон притяжения фазовой траектории. Это так называемые "странные" аттракторы, с одним из примеров такого странного аттрактора — "намоткой на тор" — мы уже познакомились в связи с некоторыми спектральными особенностями в спектре голосового источника, с другим примером аттрактора — речевым ритмом (или суперциклом), мы встречаемся на уровне управления речеобразованием [6].

A Nonlinear Model of the Acoustical Voice Source

O. P. Sklyarov

A voice source model is proposed. It takes into account the nonlinear effects of voice production by taking summation over the entire equidistant spectral series. It is shown that speech accents and stresses are related to the specific traits in the dynamic behavior of the air flow through the vocal folds. It is found that a Hopf bifurcation takes place under the first-order fluid-dynamic theory of perturbations. The proposed model is shown to fit experimental data closely.

СПИСОК ЛИТЕРАТУРЫ

1. Фант Г. Акустическая теория речеобразования. М.: Наука, 1964. 284 с.
2. Fant G. The voice source models and performance // Proc. of the XIII Int. Con. of Phonetic Sci. (ICPhS 95). Stockholm, Sweden, 13–19 Aug., 1995 / Eds. K. Klenins & P. Brandemnd. V. 3. P. 82–89.
3. Сорокин В.Н. Теория речеобразования. М., Радио и связь, 1985. 312 с.
4. Story B.H., Titze I.R. Voice simulation with a body-cover model of the vocal folds // JASA, 1995. V. 97. № 2. P. 1249–1260.
5. Скляров О.П. Динамическая модель управлением речеобразованием как процессом самоорганизации // Матер. XV Съезда оториноларингологов, 25–29 сент. 1995 / Под ред. А.А. Ланцова. Т. 1. С. 505–514.
6. Sklyarov O.P. The bifurcation model of the speech rhythm and stuttering. 4th Speech production seminar. Autrans, France. May 21–24. 1996. P. 89–93.
7. Sluijter A.M.C., Shattuck-Hufnagel S., Stevens K.N., van Heuven V. Supralaryngeal resonance and glottal pulse shape as correlates of stress and accent in English // ICPhS 95 / V. 2. P. 630–633.
8. Фланаган Д.Л. Анализ, синтез и восприятие речи. М., Связь, 1968. 392 с.
9. Beckman M.E., Trickson D., Honda K. et al. Physiological correlates of global and local pitch range variation in the production of high tones in English // ICPhS 95 / V. 2. P. 638–641.
10. Ландау Л.Д., Лифшиц Е.М. Теоретическая физика. Т. VI. Гидродинамика. М.: Наука, 1988. 733 с.
11. Шустер Г. Детерминированный хаос. М.: Мир, 1988. 237 с.