Информационные процессы, Том 7, №4, стр. 386-400.

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

К АНАЛИЗУ РЕЗОНАНСНЫХ ЧАСТОТ РЕЧЕВОГО ТРАКТА

А.С.Леонов*, В.Н.Сорокин**

** Московский инженерно-физический институт, Москва, Россия*
*** Институт проблем передачи информации, Российская академия наук, Москва, Россия*

Поступила в редколлегию 09.2007

Аннотация - Предложен метод мгновенной оценки резонансных частот речевого тракта на каждом периоде основного тона. Метод основан на нахождении интервалов времени между моментами обращения в нуль первой производной сигнала, который подвергнут фильтрации в некоторой частотной полосе, где ожидается присутствие колебаний только одного резонанса тракта. Проведены численные эксперименты, показывающие, что получаемые по этому методу оценки частот достаточно близки к истинным резонансным частотам тракта.

Применение метода к речевым сигналам для мужского голоса при многократном произнесении как изолированных гласных с различной частотой основного тона, так и гласных в симметричных слогах ГСГ с согласными /*Б*, *Г*/, показало существование частотных модуляций, доходящих до ±18%, причем фаза этих модуляций относительно момента возбуждения акустических колебаний голосовым источником изменяется в значительных пределах.

1. Введение

Проблеме определения резонансных частот речевого тракта по речевому сигналу посвящено огромное количество работ. Однако, до сих пор не существует алгоритма, который бы с высокой степенью надежности находил эти частоты. Трудности здесь связаны как с искажениями речевого сигнала при регистрации вследствие неидеальности канала связи, так и с нестационарностью процесса речеобразования. Все известные методы можно условно разделить на спектральные и временные.

Временной анализ обычно выполняется методом линейного предсказания, впервые описанного в [1]. Развито множество алгоритмов такого рода, но ни один из них не гарантирует достаточной точности определения формантных частот. Основная проблема здесь связана с тем, какое количество коэффициентов линейного предсказания следует взять. Обычно считается, что при оценке частот полюсов передаточной функции речевого тракта количество определяемых коэффициентов k_l должно определяться формулой: $k_l = N_f + C_l$, где N_f равно частоте отсчетов сигнала в кГц, а C_l - некоторая константа. Однако, было установлено, что для разных типов приемников речевых сигналов, разных условий в акустическом канале и даже на разных участках речевого сигнала в одном и том же фонетическом фрагменте число коэффициентов линейного предсказания влияет не только на точность анализа, но и на количество обнаруженных полюсов. Точность определения формантных частот методом линейного предсказания не превышает 10 – 15% [2].

Другой вариант временного анализа опирается на свойства периодичности акустических колебаний (см., например, [3]). Здесь было обнаружено влияние длительности интервала вынужденных колебаний в речевом тракте на точность оценки формантных

частот. Погрешность метода также оказалась порядка 10%, причем ошибки быстро возрастают в диапазоне частот выше 1500 Гц.

При спектральном анализе отыскиваются максимумы спектрального профиля, преобразованного с использованием аналога принципа латерального торможения, с помощью которого обостряются пики спектрального профиля [4]. Основная трудность при использовании этого метода состоит в необходимости фильтрации ложных пиков, не соответствующих формантам.

Таким образом, складывается впечатление, что анализ формантных частот должен параллельно выполняться алгоритмами, опирающимися на разные свойства речевого сигнала, а окончательный выбор частот должен определяться путем сопоставления результатов работы каждого алгоритма.

Почти все известные методы оценки резонансных частот речевого тракта по речевому сигналу основаны на предположении квазистационарности, т.е. на предположении, что резонансные частоты изменяются сравнительно медленно - со скоростями, близкими к скорости движения артикуляторных органов. Обычно принимается, что резонансные частоты остаются неизменными на интервале времени в несколько миллисекунд (см., например, работы [5 - 9]). Справедливость предположения о квазистационарности была проверена для многих методов в работе [10]. Там обнаружено, что это предположение справедливо только для синтезированного одноформантного сигнала, и только в том случае, когда эффективная длительность анализируемого речевого сегмента не больше, чем один период основного тона.

Теоретический анализ и моделирование взаимодействия между голосовым источником и речевым трактом приводят к заключению, что на интервале открытой голосовой щели резонансные частоты и их затухания отличны от их значений на интервале закрытой голосовой щели [11 - 16]. Поэтому считается, что только частоты колебательных компонент речевого сигнала, измеренные на интервале закрытой голосовой щели, соответствуют резонансным частотам речевого тракта. Но определение временного интервала закрытой голосовой щели по экспериментальным данным само по себе является трудной задачей. Во многих случаях вообще такого интервала не существует [17, 18]. Кроме того, вертикальные колебания голосовых складок создают так называемое поршневое возбуждение даже на интервале закрытой голосовой щели [16]. Поэтому методы, которые используют предположение о существовании свободных акустических колебаний на интервале закрытой голосовой щели, либо не работают, либо дают очень грубые оценки резонансных частот.

Таким образом, предположение о квазистационарности речевого сигнала и трудности нахождения подходящего сегмента для оценки формантных частот препятствуют использованию известных методов частотного анализа в обратной задаче для речевого тракта.

“Мгновенные” периоды акустических колебаний могли бы быть определены как интервалы между обращениями в нуль речевых сигналов в определенных частотных полосах. Это метод очень прост в реализации, и он интенсивно исследовался в начальный период разработок систем автоматического распознавания. К сожалению, выяснилось, что он неустойчив к шумам и смещению постоянной составляющей речевого сигнала.

Указанные трудности традиционных методов формантного анализа приводят к необходимости разработки новых надежных алгоритмов оценки параметров частотно-модулированных компонент речевого сигнала.

2. Анализ резонансных частот.

В работе [3] описан метод вычисления быстро меняющихся резонансных частот, использующий короткие по сравнению с периодом основного тона сегменты речи.

Предположим, что $y(t)$ - речевой сигнал, который отфильтрован полосовым фильтром так, что в рассматриваемой полосе содержится колебание только одной частоты. Тогда в качестве критерия оценки соответствующего периода колебаний можно использовать функцию

$$Q_0(t) = \frac{y''(t)}{y'(t) - \overline{y}'(t)}. \qquad (1)$$

Здесь $y'(t)$ и $y''(t)$ - производные по времени, а $\overline{y}'(t)$ - производная, сглаженная с прямоугольным окном на интервале, примерно равном ожидаемому периоду основного тона. В таком критерии подавляются постоянная составляющая речевого сигнала и его шумовые составляющие. Фильтрация шумов обеспечивает также устойчивость численного дифференцирования в формуле (1).

Оказалось, что для свободных колебаний с периодом T интервалы времени Δt между точками разрыва функции $Q_0(t)$ достаточно близки к величине $T/2$, и это позволяет вычислять оценки частот этих колебаний как $F = 1/(2\Delta t)$. Экспериментально было найдено, что небольшие вариации длительности окна сглаживания мало влияет на оценку частот.

В данной работе оценки формантных частот проводятся с использованием более простого и естественного критерия:

$$Q(t) = y'(t). \qquad (2)$$

Речевой сигнал, представленный в виде отсчетов в моменты времени t_k, следующие с шагом h_s, анализируется на скользящем сегменте длительностью T_0. Шаг смещения сегмента равен h_s.

Алгоритм оценивания формантных частот состоит из следующих шагов.

1) Речевой сигнал фильтруется гребенкой цифровых фильтров с амплитудно-частотной характеристикой $W_i(f)$, выбранной таким образом, чтобы в заданной области частот $f_{i\min} < f < f_{i\max}$, предположительно, находилась только одна форманта.
2) Сигнал в каждой частотной полосе дифференцируется, т.е. величина (2) вычисляется конечно-разностным методом на сетке значений $\{t_k\}$.
3) Отыскиваются отрезки времени $[t_{k(n)}, t_{k(n+1)}]$, содержащие переходы через нуль функции (2).
4) Выполняется линейная интерполяция функции (2) на найденных отрезках и приближенно оценивается момент $\tau_n \in [t_{k(n)}, t_{k(n+1)}]$ перехода функции (2) через нуль.
5) На текущем сегменте речевого сигнала вычисляются все возможные предварительные оценки формантных частот как $F_n = 1/(2(\tau_{n+1} - \tau_n))$.
6) В каждой частотной полосе строится гистограмма распределения этих оценок.
7) Частота $\tilde{F}$, которая соответствует максимуму гистограммы, принимается как оценка форманты из выбранного частотного диапазона $f_{\min} < f < f_{\max}$.

Покажем на простом примере свободных затухающих колебаний в речевом тракте с закрытой голосовой щелью, что такая оценка дает адекватные результаты. Примем следующую грубую модель «отфильтрованного» сигнала. Пусть

$$y(t) = Ae^{-gt}\sin(pt+\varphi) + B,\ p^2 = (2\pi)^2(F^2 - G^2),$$

где A, φ – амплитуда и фаза колебаний, B - постоянная составляющая сигнала, F - частота свободных колебаний, g – декремент затухания, определяемый потерями в тракте, а $G = g/(2\pi)$. Нетрудно найти нули τ_n $(n = 1,2,...)$ функции (2) для такого сигнала. Они определяются уравнением

$$\operatorname{ctg}(p\tau_n + \varphi) = \frac{g}{p}.$$

Используя неравенство $|x - x_n| \le |\operatorname{ctg} x|$, где $x_n = \pi/2 + \pi(n-1)$, и которое справедливо для $|x - x_n| < \pi/2$, можно найти, что

$$\pi - 2\frac{g}{p} \le p(\tau_{n+1} - \tau_n) \le \pi + 2\frac{g}{p},$$

Отсюда следует, что оценки F_n из п. 5 алгоритма для формантной частоты F заключены в пределах:

$$\frac{\sqrt{F^2 - G^2}}{1 + \dfrac{2G}{\pi\sqrt{F^2 - G^2}}} \le F_n = \frac{1}{2(\tau_{n+1} - \tau_n)} \le \frac{\sqrt{F^2 - G^2}}{1 - \dfrac{2G}{\pi\sqrt{F^2 - G^2}}}.$$

Из этих неравенств можно найти относительные погрешности оценок F_n, которые дает алгоритм:

$$\delta_F = \frac{|F_n - F|}{F} \le \frac{4}{\pi}\frac{G}{F}\frac{[1 - (G/F)^2]}{[1 - (1 + 4\pi^{-2})(G/F)^2]}.$$

Это также верно и для итоговой оценки форманты $\tilde{F}$. Таким образом, при «малых потерях», т.е. при $G/F \ll 1$ оказывается, что и погрешность определения форманты тоже мала: $\delta_F \le 1.3(G/F)$.

Подобные аналитические исследования алгоритма в случае более сложных сигналов затруднены, даже если голосовая щель тракта считается закрытой. Поэтому для более реалистичной модели колебаний тракта мы проведем численное исследование алгоритма 1) - 7). Будем считать, что ищется формантная частота F и известны ее границы: $f_{\min} < F < f_{\max}$. Предположим также, что отфильтрованный речевой сигнал имеет вид волнового пакета

$$y(t) = \sum_{j=1}^{2m+1} W_j y_j(t), \qquad (3)$$

т.е. суперпозиции колебаний $y_j(t)$ одиночных не связанных резонаторов, которые описываются дифференциальными уравнениями вида:

$$y''+2g_j y'+(2\pi f_j)^2 y = G(t),\quad t_0 < t < T_0. \qquad (4)$$

Здесь $f_j \in (f_{\min}, f_{\max})$ - собственная частота *j*–го резонатора, W_j - вес колебаний *j*–го резонатора в волновом пакете, определяемый окном фильтрации $W(f)$, g_j - декремент затухания для *j*–го резонатора, $G(t)$ - квазипериодическое возбуждение голосового источника со средним периодом t_0, а T_0 – длина речевого сегмента, используемого для оценки форманты.

Начальные условия для уравнений (4) задаются в два этапа. Сначала уравнения (4) решаются нулевыми начальными условиями на отрезке первого периода голосового источника $[0, t_0]$. Полученные в результате этого величины $y_j(t_0), y_j'(t_0)$ принимаются как новые начальные условия для уравнений (4). Далее выписывается аналитические решения уравнений (4) с этими начальными условиями:

$$y_j(t) = A_j e^{-g_j t}\sin(p_j t+\varphi_j) + \frac{1}{p_j}\int_0^t e^{-g_j(t-\tau)}\sin p_j(t-\tau)G(\tau)d\tau\ ,$$

где A_j, φ_j – амплитуда и фаза *j*–го колебания, определяемые начальными условиями, а $p_j = \sqrt{(2\pi f_j)^2 - g_j{}^2}$. Эти решения позволяют вычислить отфильтрованный сигнал (3). Такая модель сигнала учитывает как собственные колебания тракта, так и его вынужденные колебания при открытой голосовой щели. Однако, влияние голосовой щели на собственные частоты речевого тракта при этом не учитывается.

Используем эту модель отфильтрованного сигнала для численной оценки формант по предлагаемому алгоритму. Уточним детали. Численные эксперименты проводились для различных величин F из диапазона от 0.3 Кгц до 3.3 Кгц. Временная сетка сигнала (3) определялась частотой отсчетов 16 Кгц. Частоты f_j, входящие в волновой пакет, задавались в виде

$$f_j = F + (j-m-1)(f_{\max}-f_{\min})/(2m),\ \ j = 1,2,...,2m+1$$

с помощью априорно известных границ форманты $f_{\min}, f_{\max}$ и с «центральной частотой» $F = (f_{\min}+f_{\max})/2$ для $m = 15$. Весовые коэффициенты W_j задавались как в известном окне фильтрации Хемминга. Для определения декрементов затухания g_j, а также полуширины окна Хемминга, использовалась известная зависимость $g = g(f)$ из книги [16], показанная на рис. 2.

Голосовой источник имел средний период 8 мс и моделировался функцией $G(t)$ из [16]. Ее вид дан на рис. 1(1). На рис. 1(2), 1(3) показан модельный отфильтрованный сигнал (3) для случая $F = 1.51$ Кгц и его производная, вычисленная на временной сетке по формуле $y'(t_k) \approx 0.5[y(t_{k+1}) - y(t_k)]/h_s$. Красными точками показаны приближенно найденные нули этой производной (см. п.п. 3, 4 алгоритма), которые используются в п. 5 для нахождения «мгновенных» оценок форманты. На рис. 1(4) приведена временная зависимость этих оценок (кружки) в сопоставлении с искомой величиной F (красная линия). Для сравнения на рис. 1(5) дан график зависимости площади голосовой щели от времени. Из этих двух последних

графиков ясно, что наилучшие и стабильные оценки формант получаются именно для закрытой голосовой щели.

На рис. 3 представлено распределение мгновенных оценок форманты в виде гистограммы для $F = 1.51$ Кгц. Там же звездочкой показано положение итоговой оценки $\tilde{F}$ формантной частоты. Кружком показано среднее значение мгновенных оценок. Аналогичные гистограммы для $F = 0.51$ Кгц и $F = 2.51$ Кгц показаны на рис.4, 5.

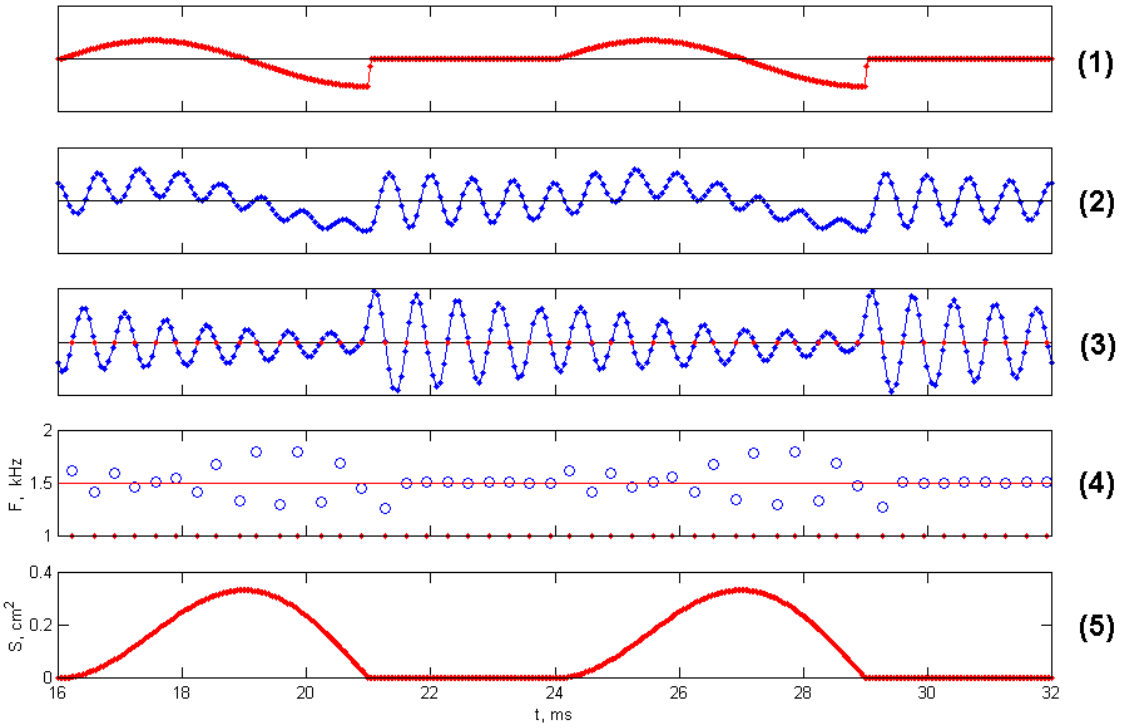

Рис. 1

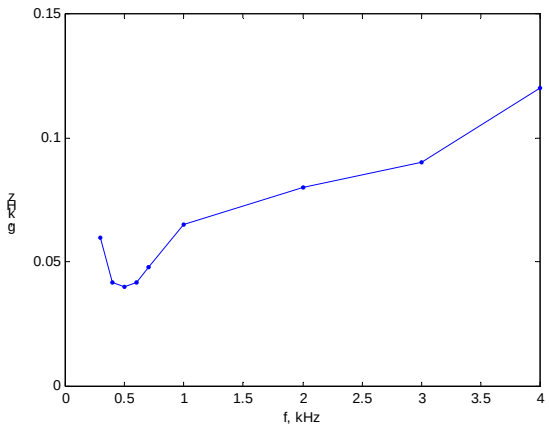

Рис.2. Декремент затухания.

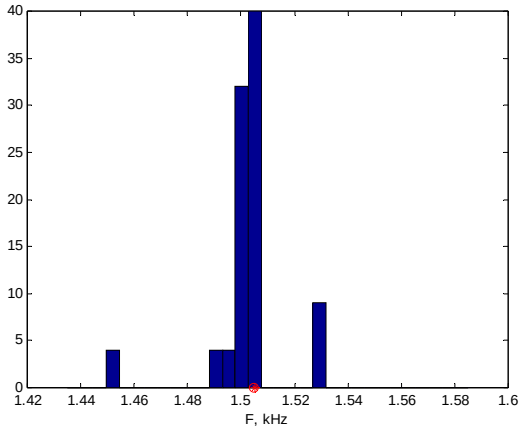

Рис.3. $F = 1.51$ Кгц

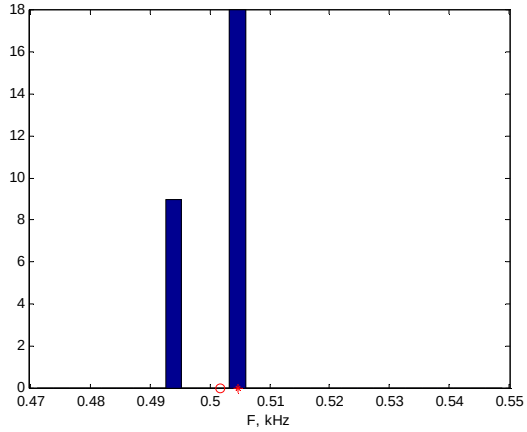

Рис.4. $F = 0.51$ Кгц

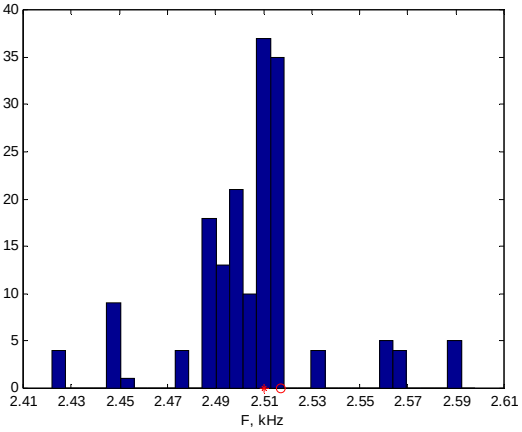

Рис.5. $F = 2.51$ Кгц

Для различных модельных формантных частот F из указанного выше диапазона были найдены относительные ошибки $\delta_F = |\tilde{F} - F| / F$ оценок, полученных по алгоритму. График зависимости этих ошибок от F приведен на рис. 6. Таким образом, численные эксперименты с модельным отфильтрованным сигналом показывают, что алгоритм дает оценки формант с точностью не хуже, чем 3 %. Отметим также, что среднее значение мгновенных оценок формант приближает истинное значение хуже, чем оценка по гистограмме.

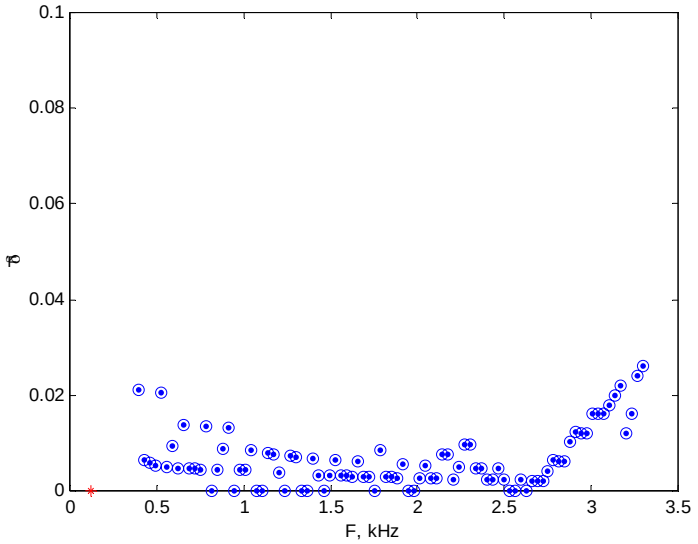

Рис.6. Относительные погрешности оценок алгоритма.

Дальнейшие численные эксперименты по верификации предложенного алгоритма гистограмм связаны с его проверкой для простых реальных сигналов, у которых формантные частоты хорошо известны. В качестве таковых были выбраны звуковые сегменты английских гласных из слов «but» (в произношении мужчины и женщины), «beet», «bat» и «foot» (в произношении мужчины). Использованы экспериментальные данные с частотой отсчетов 16 Кгц. Результаты сведены в Таблицу 1, где представлены полученные оценки формант и их относительные ошибки.

Таблица 1. Резонансные частоты гласных (кГц) и относительные ошибки их оценивания.

Звук	1-я форманта	2-я форманта	3-я форманта	Ошибка 1-й форманты	Ошибка 2-й форманты	Ошибка 2-й форманты
but_man	0.622	1.112	2.070	0.021	0.038	0.033
but_woman	0.770	1.390	2.365	0.013	0.021	0.002
beet_man	0.409	2.137	2.777	0.022	0.028	0.008
bat_man	0.648	1.868	2.491	0.063	0.002	0.011
foot_man	0.550	1.053	2.469	0.016	0.008	0.038

Из Таблицы 1 видно, что точность определения формант по алгоритму гистограмм оказывается для реального сигнала несколько ниже, чем для модельного. Однако, эта точность не хуже 6.5%, а в среднем составляет около 2%.

Применяя алгоритм гистограмм для сегментов исследуемого речевого сигнала длиной 1 – 5 периодов звукового источника, можно получить оценки динамических треков формантных частот. Об этом речь пойдет в следующих разделах.

3. Эксперименты с динамическими сегментами речи. Модуляции формант.

В этом разделе используются реальные звуковые сигналы с несколько иными характеристиками, чем в предыдущем. Например, выделение частотной области с предполагаемым положением единственной форманты выполнялось с помощью гребенки фильтров Баттерворта 10-го порядка с крутизной склонов около 60 дБ. Речевой сигнал дискретизировался с частотой отсчетов 20 кГц, т.е. с периодом в 50 мкс. Поэтому погрешность дискретизации при определения периода формантной частоты в диапазоне от 200 Гц до 2000 Гц составляет 0.5 – 5%. В частности, для частоты 600 Гц, характерной для первой форманты гласного /*А*/, ошибка вследствие дискретизации составляет 1.5%.

Речевой сигнал принимался через направленный микрофон, обеспечивающий отношение сигнал/шум 60 дБ, и квантовался на 16 бит стандартным преобразователем в персональном компьютере.

Все эксперименты проводились на речевом материале, записанном от одного диктора мужского пола. Речевой материал состоял из отдельно произнесенных гласных русского языка, а также симметричных слогов ГСГ с согласными /*Б*/ и /*Г*/ и ударением на первом согласном. Каждый изолированный гласный произносился по 5 раз, а каждый слог произносился по 10 раз. Как изолированные гласные, так и гласные в слогах произносилась с разной частотой основного тона в диапазоне 90 – 180 Гц.

Для этого экспериментального материала предложенный выше алгоритм не только дает оценки формантных частот, близкие к оценкам, полученным другими методами, но и обнаруживает некоторые особенности речевых сигналов, которые в других методах не регистрируются. На рис. 7 показаны оценки периодов первой и второй форманты в слоге /*ЭБЭ*/. Здесь, прежде всего, следует обратить внимание на устойчивую оценку периода первой форманты на звонкой смычке, и на правдоподобную динамику этой форманты в окрестностях смычки. Как известно, на интервале звонкой смычки первый резонанс речевого тракта не стремится к нулю, как это было бы в акустической трубе с абсолютно жесткими стенками. Вместо этого в речевом тракте присутствует резонанс радиальных колебаний F_r, и эти колебания излучаются через щеки и мягкие ткани шеи. В зависимости от объема речевого тракта и характеристик тканей, частота этого резонанса находится в диапазоне 150 – 350 Гц (периоды колебаний 6.7 – 2.8 мс). Поскольку эти частоты близки к частотам основного тона, то оценка F_r обычно затруднительна.

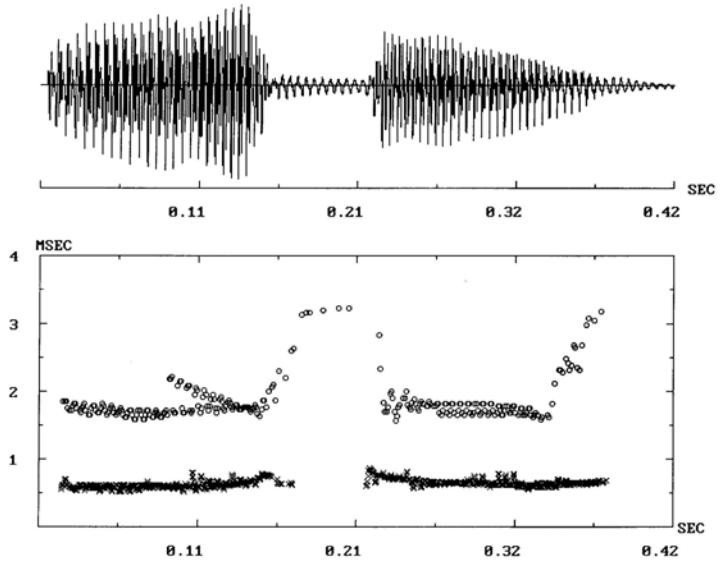

Рис. 7. Оценки периодов первой (○) и второй (x) форманты в слоге /*ЭБЭ*/

На рис. 8 показано распределение оценок частоты радиального резонанса по множеству слогов с согласными /*Б*/ и /*Г*/, произнесенными с разной частотой основного тона. Из этого распределения видно, что нет явной зависимости между частотой радиального резонанса F_r и частотой основного тона F_0. Вместе с тем, оценки F_r для смычки /*Б*/ расположены ниже оценок для /*Г*/. Это соответствует свойствам радиального резонанса при смычке в разных местах речевого тракта. При смычке на губах объем речевого тракта больше объема при смычке в области мягкого неба. Общая податливость стенок для /*Б*/ также больше, чем для /*Г*/ из-за колебаний щек. Поэтому частота радиального резонанса для /*Б*/ должна быть ниже, чем для смычки /*Г*/.

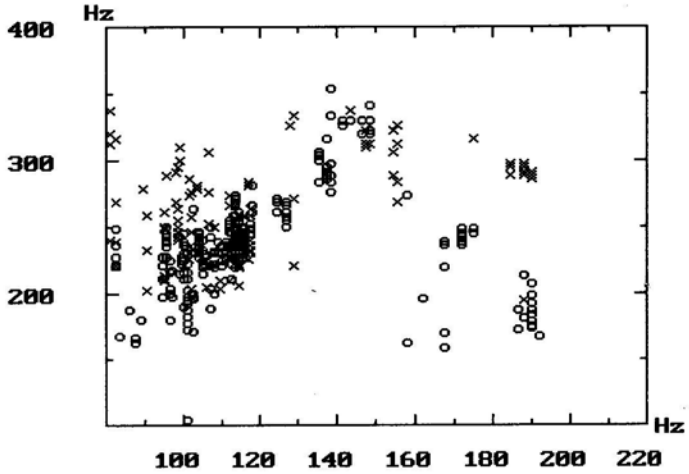

Рис. 8. Оценка частоты радиального резонанса на смычке согласных /*Б*/ (○) и /*Г*/ (x) в зависимости от частоты основного тона.

Оценка F_r на интервале звонкой смычки создает дополнительный признак для различения губных и заднеязычных согласных в процедурах автоматического распознавания речи.

На рис. 7 видно, что оценки периодов формантных частот как бы колеблются около некоторого стабильного положения. При этом интересно появление колебаний оценок какого-то дополнительного резонанса в низкочастотной области на интервале времени от

0.11 сек до начала смычки. Такие резонансы часто наблюдаются в речевом сигнале, и их интерпретация довольно затруднительна, в том числе и по причине нестабильной оценки частот.

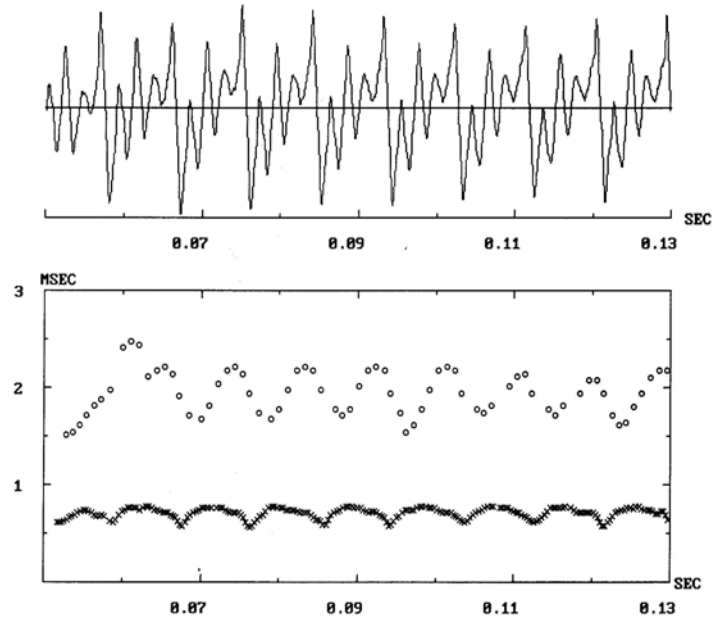

Рис. 9. Модуляции периодов первой (○) и второй (x) форманты гласного /*Э*/ .

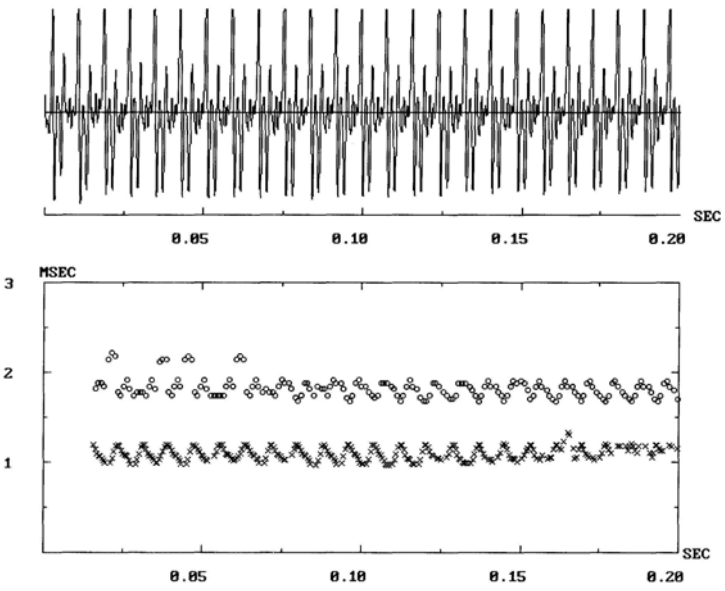

Рис. 10. Модуляции периодов первой (○) и второй (x) форманты гласного /*О*/.

В проведенных численных экспериментах выявились значительные частотные модуляции внутри периода основного тона почти для всех гласных. Исключение составляет частота первой форманты для гласных /*Ы,И*/. На рис. 9, 10 видны модуляции как первой, так второй формантной частоты у гласных /*Э,О*/, синхронные с основным тоном.

Если определить индекс модуляции δf как максимальное отклонение частоты форманты от ее среднего значения, то для первой форманты гласного /*А*/ $\delta f_1 = \pm 5.3\%$ (среднее значение F_1=651 Гц), а для второй форманты со средним значением F_2=1132 Гц

δf_2=±10.5%. Соответственно, для гласного /Э/, δf_1= ±6.1% и δf_2= ±8% при средних значениях формантных частот F_1=553 Гц и F_2=1680 Гц. Индексы модуляции для гласного /*О*/ равны δf_1= ±11.4% и δf_2= ±*5*.8%. Диапазон максимальных и минимальных значений формантных частот приведен в Таблице 2.

Индексы частотной модуляции первой и второй формант мужского голоса практически не зависят от частоты основного тона, демонстрируя лишь слабое уменьшение в диапазоне 130 – 180 Гц.

Таблица 2. Модуляции первой и второй формантной частоты (Гц).

Гласная	$F_{1\ min}$	$F_{1\ max}$	$F_{2\ min}$	$F_{2\ max}$
А	600	732	952	1333
Э	454	536	1304	1714
О	483	588	815	1034
У	361	526	750	780
Ы	250	268	1714	2000
И	224	226	1764	2400

Наблюдается также значительное разнообразие как амплитуд частотных модуляций, так и их фаз относительно момента голосового возбуждения. Например, для гласного /*А*/ максимальное значение периода первой форманты (наименьшее значение частоты на периоде модуляции) сдвинуто по времени примерно на 40% периода основного тона относительно момента возбуждения, тогда как минимум периода форманты (наибольшее значение частоты форманты на периоде модуляции) обычно совпадает моментом голосового возбуждения. Однако, при другом произнесении этого гласного минимум периода первой форманты форманты оказался сдвинут относительно момента возбуждения.

В противоположность гласному /*А*/, на гласном /Э/, на 40% от периода основного тона оказался сдвинут не максимум периода первой форманты, а его минимум. Иногда на одном и том же произнесении наблюдается сдвиг фазы частотной модуляции или резкое изменение индекса модуляции.

Сдвиг фазы модуляций иллюстрируется рис. 11 – 13 для одного и того же гласного /*А*/, произнесенного с разной частотой основного тона.

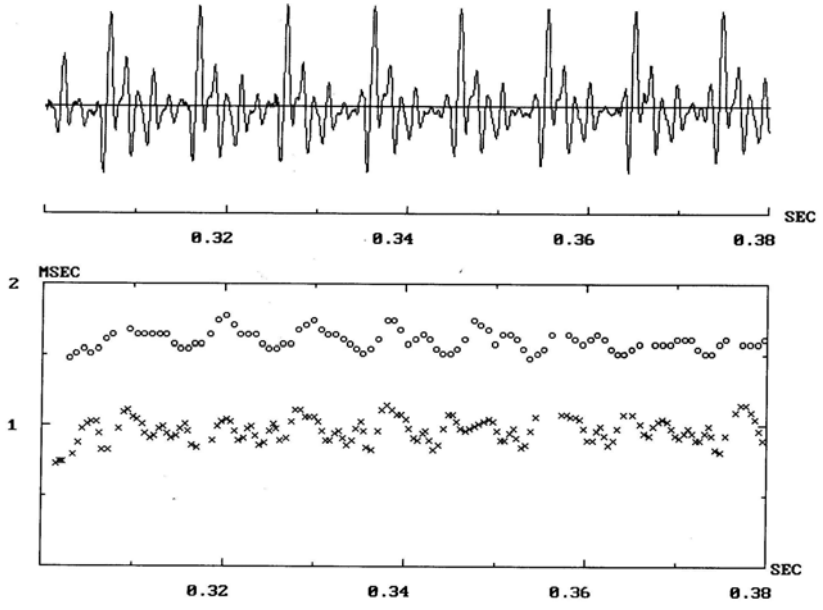

Рис. 11. Модуляции периодов первой (○) и второй (х) форманты гласного /*А*/ при

частоте основного тона F_0=102 Гц.

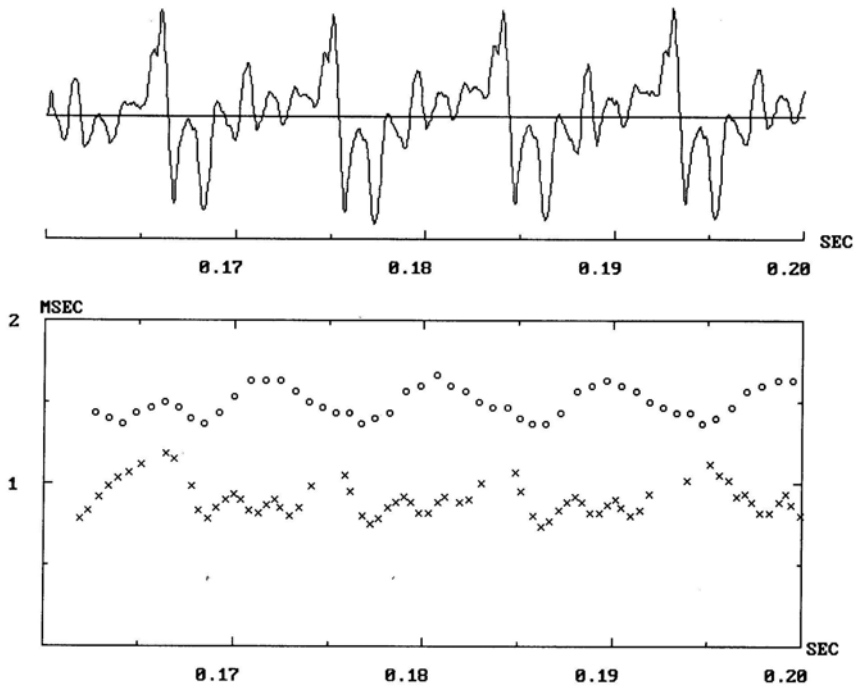

Рис. 12. Модуляции периодов первой (○) и второй (x) форманты гласного /*A*/ при частоте основного тона F_0=110 Гц.

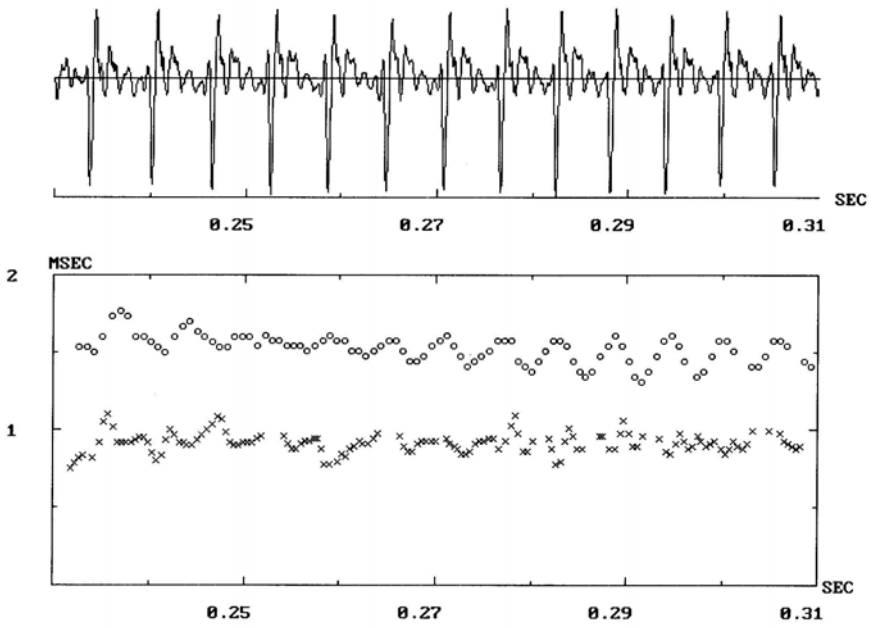

Рис. 13. Модуляции периодов первой (○) и второй (x) форманты гласного /*A*/ при переходе частоты основного тона F_0 от 123 Гц до 165 Гц..

4. Обсуждение

Предложенный в данной работе алгоритм оценки мгновенных формантных частот в применении к реальным сигналам показывает существование временных модуляций формант в виде соответствующих значимых колебаний оценок. Явление модуляции формант хорошо известно. Это экспериментально установленное свойство речевого сигнала. Однако, иногда его связывают с эффектами дискретизации речевого сигнала или с недостатками конкретного алгоритма вычисления формант. Вследствие высокой точности используемого в данной работе алгоритма, можно утверждать, что не в этом дело и что эффект модуляции формант связан с объективно существующими динамическими процессами в акустике речеобразования.

Причины модуляций формант скорее всего многообразны и не связаны с каким-либо единственным механизмом. Один из таких предполагаемых механизмов состоит в параметрическом изменении резонансных частот речевого тракта вследствие изменения граничных условий при открытой голосовой щели. При этом не только могут меняться резонансные частоты речевого тракта, но в речевом сигнале могут появляться акустические колебания, частота которых определяется свойствами подсвязочной области – трахеи, бронхов и легких.

Различные модели взаимодействия речевого тракта и подсвязочной области подтверждают возможность возникновения частотных модуляций. Так, в работе [12] была выполнена оценка изменения резонансов тракта при открытии голосовой щели (Таблица 3).

Таблица 3. Вариации первой резонансной частоты (Гц) по [12].

Гласная	Закрытая голосовая щель	Площадь голосовой щели 0.08 $см^2$	Площадь голосовой щели 0.12 $см^2$
A	677	806	858
E	459	475	482
O	538	582	582
U	291	308	323
I	285	297	305

Несколько иная модель взаимодействия была описана в [14] (Таблица 4).

Таблица 4. Вариации первой резонансной частоты (Гц) по [14].

Гласная	Закрытая голосовая щель	Площадь голосовой щели 0.027 $см^2$
A	676.1	682.8
E	435.0	436.2
O	535.1	540.4
U	248.8	249.5
I	230.1	230.5

Работе [16] было показано, что, при определенном соотношении импеданса речевого тракта и подсвязочной области, и с учетом переменной скорости звука в голосовой щели, частота первого резонанса однородной акустической трубы увеличивается на 9.2% при открытой голосовой щели. Однако, там же было установлено, что знак частотной модуляции может смениться на обратный при определенных условиях.

Разнообразные динамические явления могут повлиять не только на амплитуду частотных модуляций, но и на сдвиг фазы относительно момента смыкания голосовых

складок. Например, для речевого тракта длиной в 17.5 см при скорости звука 350м/с акустическая волна распространяется от голосовой щели до губ и обратно за 1 мс. За это время площадь открытой голосовой щели может заметно измениться, и это изменение зависит от периода основного тона. При частоте основного тона 130 Гц, характерной для мужских голосов, период основного тона составляет около 7.6 мс. Длительность интервала открытой голосовой щели обычно составляет около 0.6 от периода основного тона, т.е., в данном случае, примерно 4.6 мс, что сопоставимо со временем распространения волны в речевом тракте. Соотношение между интервалом открытой голосовой щели и временем распространения акустической волны у женских голосов со средней частотой основного тона около 250 Гц еще меньше. Существуют и другие динамические эффекты, которые могут повлиять на частотные характеристики речевого тракта.

Максимальная скорость воздушного потока на выходе из голосовой щели может доходить до 30 – 35 м/с, что составляет заметную долю от скорости звука. Переменная скорость воздушного потока изменяет скорость распространения акустической волны от губ до голосовой щели, а с ней и значения резонансных частот речевого тракта. Теоретический анализ [16] и эксперименты на физических моделях [17, 18, 19] показывают, что на выходе из голосовой щели возникает турбулентность воздушного потока, на которой возможно частичное отражение звуковых волн. Эта турбулентность продолжает существовать и при сомкнутых голосовых складках [20, 21], так что взаимодействие звуковых волн с турбулентным потоком продолжается и на этом интервале периода основного тона с возможным изменением акустических характеристик речевого тракта.

5. Заключение

Разработанный алгоритм частотно-временного анализа речевого сигнала позволяет получить сравнительно точные оценки мгновенных резонансных частот речевого тракта. Численные эксперименты по оценке мгновенных значений периодов первых двух резонансных частот речевого тракта показывают существование значительных частотных модуляций на периоде основного тона. Фаза этих модуляций не фиксирована относительно момента смыкания голосовых складок, поэтому определение формантных частот на интервале закрытой голосовой щели может сопровождаться значительными ошибками.

Этот алгоритм создает новые возможности для решения таких прикладных задач, как идентификация и верификация диктора, распознавание речи и сжатие речевого сигнала в каналах связи.

Литература

1. B.S.Atal, S.L.Hanauer. Speech analysis and synthesis by linear prediction of the speech wave, J. Acoust. Soc. Am., 1971, v. 50, pp. 637-655.
2. G.K.Vallabha, B.Tuller. Systematic errors in formant analysis of steady-state vowels. Speech Communication, 2002, v. 38. pp. 141-160.
3. В.Н.Сорокин, И.П.Трифоненков. Об автокорреляционном анализе речевых сигналов, Акуст. ж., 1996, т.42, № 3, 368-374.
4. В.Н.Сорокин, А.И.Цыплихин. Сегментация и распознавание гласных. Информационные процессы, 2004, т.4, №2, 202-220.
5. K.S.Nathan, Y.-T.Lee, H.F.Silverman. A time-varying analysis method for rapid transitions in speech. IEEE Trans. Signal Processing, 1991, v.39, pp. 815-824.
6. P.Maragos, J.F.Kaiser, T.F.Quatieri. On amplitude and frequency demodulation using energy operators. IEEE Trans. Signal Processing, 1993a, v. 41, pp. 1532-1550.

7. P.Maragos, J.F.Kaiser, T.F.Quatieri. Energy separation in signal modulation with application to speech analysis. IEEE Trans. Signal Processing, 1993b, v. 41, pp. 3024-3051.
8. A.Potamianos, P.Maragos. Speech formant frequency and bandwidth tracking using multiband energy demodulation. J. Acoust. Soc. Am., 1996, v. 99, pp. 3795-3806.
9. S.Lu, P.C.Doerschuk. Nonlinear modeling and processing of speech based on sums of AM-FM formant models. IEEE Trans. Signal Processing, 1996, v.44, pp. 773-782.
10. R.Smits. Accuracy of quasistationary analysis of highly dynamic speech signals. J. Acoust. Soc. Am. 1996, v. 96, pp. 3401-3415.
11. J.Flanagan. Speech Analysis, Synthesis and Perception. (New York, Berlin) (1964).
12. G.Fant , H.Wakita: Toward a better vocal tract model. STL\ QPSR, 1978, Speech Transmission Laboratory, Royal Institute of Technology, Stockholm, Sweden, pp. 9-29.
13. T.V.Anantapadmanbha, G.Fant. Calculation of true glottal flow and its components. Speech Communication, 1982, v. 1, pp. 167-184.
14. P.Badin, G.Fant. Notes on vocal tract computation. STL-QPSR 2 - 3 (1984), Speech Transmission Laboratory, Royal Institute of Technology, Stockholm, Sweden, pp. 53-108.
15. P.Meyer, W.Strube. Calculations on the time varying vocal tract. Speech Communication, 1984, v. 3, pp. 109-122.
16. В.Н.Сорокин. Теория речеобразования. М., Радио и связь. 1985.
17. D.H.Klatt, L.C.Klatt: Analysis, synthesis and perception of voice quality variations among female and male talkers. J. Acoust. Soc. Am., 1990, v. 87, pp. 820-857.
18. B.Holmberg, R.T.Hillman, J.S.Perkell. Glottal air flow and pressure measurements for soft, normal and loud voice by male and female speakers. J. Acoust. Soc. Am., 1988, v. 84, pp. 511-529.
18. C.G.Guo , R.C.Scherer. Finite element simulation of glottal flow and pressure. J. Acoust. Soc. Am., 1993, v. 92, Part 2, pp. 688-700.
19. X.Pelorson, A.Hirschberg, A.P.J.Wijnands, H.Bailliet. Description of the flow through models of the glottis during phonation, Acta Acustica, 1995, v. 3, pp. 191-202.
20. X.Pelorson, A.Hirschberg, R.R. van Hass, A.P.J.Wijnands, Y.Auregan. Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. Application to a modified two-mass model. Acoust. Soc. Am., 1994, v. 99, pp. 3416-3431.
21. X.Pelorson, J.Liljencrants, B.Kroeger. On the aeroacoustics of voiced sound production. Proc. Intern. Congress on Acoustics, Trondheim , 1995, v. 4, pp. 501-504.