

## ИНДИКАЦИЯ ВЫСОТЫ ГОЛОСА ПО СИНХРОННОСТИ ИЗМЕНЕНИЯ ВЫСОТЫ ЕГО ОБЕРТОНОВ

*Д.т.н., профессор В.Р. Женило (Академия управления МВД России)*

Что такое высота голоса и как ее следует измерять – это одна из многих все еще нерешенных задач, стоящая в одном ряду со всеми другими задачами из области искусственного интеллекта.

Поначалу это утверждение может показаться неверным, поскольку все мы легко и не безошибочно умеем определять изменение высоты голоса певца или говорящего. Существует даже огромное количество аппаратных и программных методов измерения и визуализации динамики высоты голоса. Но то, что выдают такие приборы, – это результат, который невозможно проверить. Невозможно потому, что все наши оценки высоты голоса исключительно субъективны.

Хлопки голосовых связок – это отдельные импульсы, в которых изначально никаких основных тонов и обертонов нет. Это хорошо видно на широкополосных сонограммах, где след каждого голосового импульса проявляется в виде вертикальных линий. Горизонтальные же линии, которые мы и называем треками основного тона и треками обертонов, на сонограмме появляются лишь тогда, когда в окно спектрального анализа попадает не менее двух голосовых импульсов. А если окно спектрального узкополосного анализа имеет постоянную ширину, то для низких мужских голосов в него может попасть пара-тройка голосовых импульсов, а для высоких женских или детских – в 4-6 раз больше. А поскольку в натуральной речи соседние голосовые импульсы всегда несколько отличаются друг от друга, то треки обертонов голоса строго никогда не подчиняются закону – частота  $N$ -того обертона (или гармоники) строго равна частоте основного тона (основной гармоники) умноженной на  $N$ . Таким свойством обладают только строго периодические сигналы, каковых в живой природе не бывает.

Смещения обертонов реального (живого) голоса относительно идеальной целочисленной решетки, построенной на базе основного тона, будут разными, в зависимости от ряда факторов: ширины временного окна спектрального анализа, передаточной функции речевого тракта и, как ни странно, от передаточной функции слуховой системы, которая у каждого слушателя тоже имеет свои индивидуальные особенности. Видимо, поэтому бывают казусы, когда два очень опытных педагога вокалиста соглашаются с тем, что испытуемый правильно исполнил всю партитуру, за исключением одной ноты, которую, по мнению одного из педагогов, испытуемый исполнил явно неверно.

Каким образом мы реально оцениваем высоту голоса – это еще не разгаданная загадка природы. В данном докладе предлагается еще один взгляд на эту проблему.

Исследовать поведение обертонов голоса удобно с помощью узкополосных сонограмм. В данном исследовании использовалась компьютерная программа «Мастерская сигналов», в которой имеется возможность не только оценивать мгновенные частоты обертонов голоса в каждом отдельном спектральном кадре, но и, что очень важно, оценивать скорости изменения частот этих обертонов всего лишь по одному мгновенному комплексному спектру. Ранее такие оценки проводились только по паре смежных амплитудных спектров.

Пусть  $F_0$  – это частота основного тона голоса в текущем кадре спектрального узкополосного анализа речевого сигнала. Обозначим  $F_n$  частоту в точке локального максимума этого спектра в районе обертона. Принимая, как обычно, для первого приближения  $F_n = nF_0$ , мы будем иметь такое же соотношение и для производной по времени этих же частот –  $F'_n = nF'_0$ . Введем новый параметр:

$$p_n = \frac{F'_n}{F_n}.$$

Очевидно, что этот параметр оказывается независимым от номера обертона. Иначе говоря, параметр  $p_n$  инвариантен к номеру обертона. А раз так, то, построив график зависимости параметра  $p_n$  от частоты локального максимума спектра, мы должны будем получить во всех точках некоторое постоянное значение.

Действительно, на рис. 1 на заднем плане показана сонограмма искусственного сигнала, у которого все гармоники ведут себя так, как они должны себя вести теоретически в голосе, высота которого плавно нарастает.

Если построить такой же рисунок для живого речевого сигнала, то он будет испещрен множеством, казалось бы, случайно разбросанных точек, каждая из которых соответствовала бы конкретному локальному максимуму амплитудного спектра. Чтобы убрать лишние точки, явно не соответствующие обертонам голоса было сделано следующее. Во-первых, из рассмотрения были выброшены все локальные максимумы, частоты которых превышали 2 КГц. Во-вторых, была предпринята попытка описать свойство любого обертона, основанное на высокой предсказуемости появления его в соседнем спектральном кадре. А именно, в каждом текущем кадре для каждого локального максимума амплитудного спектра оценивалась скорость роста во времени частоты этого локального максимума. И на основании этой оценки получались предсказания – где этот локальный макси-

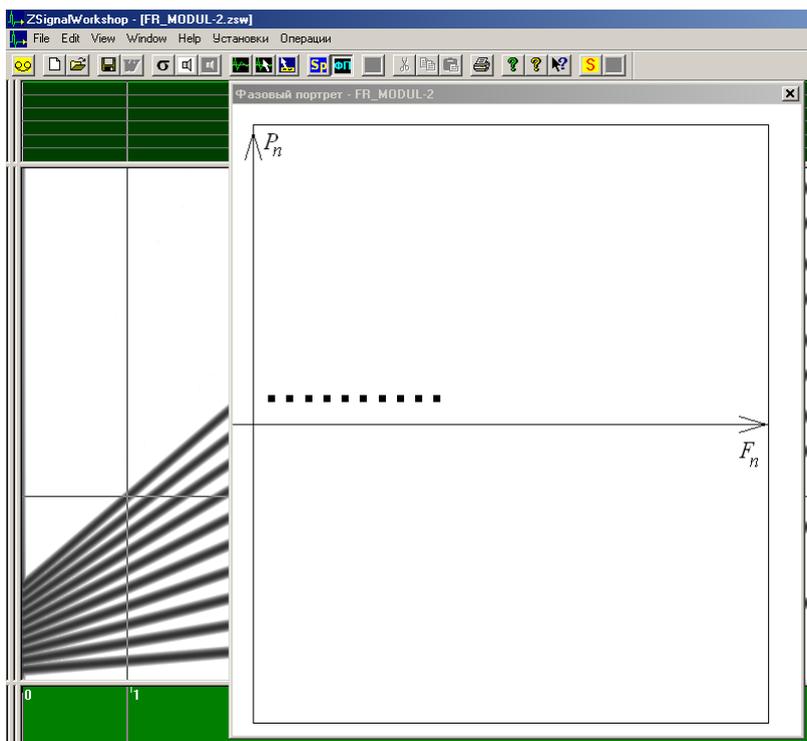


Рис. 1. Синхронность изменения частот гармонических сигналов.

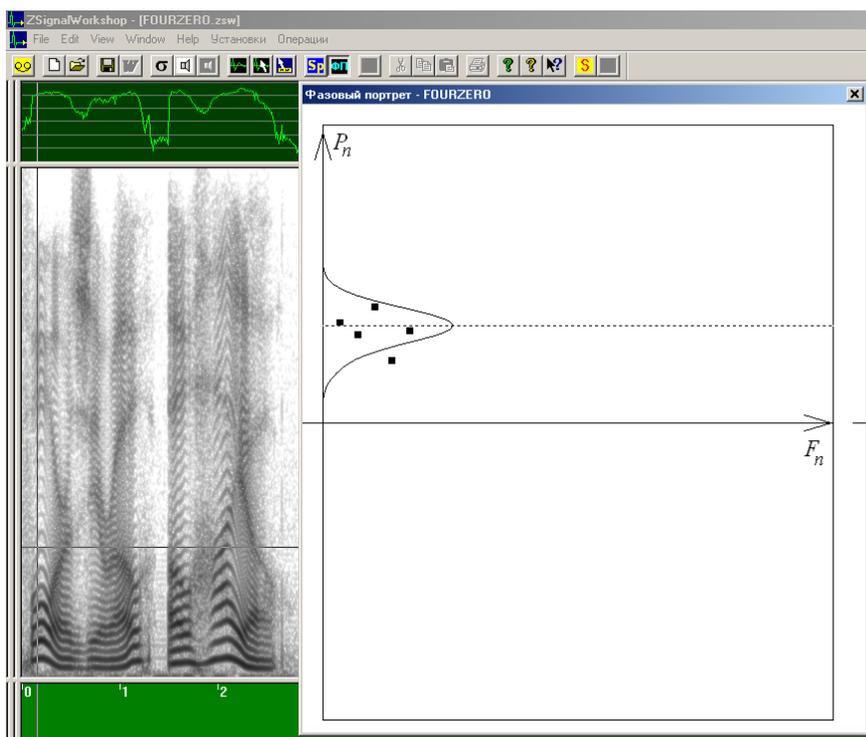


Рис. 2. Нарушение синхронности роста частот обертонов голоса.

качество звучания. Поэтому можно утверждать, что описанное явление свойственно именно самому речевому сигналу. Хотя, следует отметить, что повторное воспроизведение и запись этого же речевого сигнала в обычном помещении привело к дополнению эффекта реверберации, что еще сильнее размазало распределение параметра  $p_n$  на рис. 2 – 8.

мум должен появиться в смежных (предыдущем и последующем) кадрах. Расхождения предсказанных с реальными значениями принимались за меру ошибки предсказания<sup>1</sup>.

На основе ошибок предсказания и общих ошибок всех измеряемых физических величин, для всей совокупности вычисленных значений параметров  $p_n$  вычислялась функция плотности вероятности распределения параметров  $p_n$  в данном спектральном кадре. Если бы все обертона голоса изменялись синхронно, как, например, в искусственном сигнале на рис. 1, то такая функция плотности вероятности имела бы вид функции Гаусса, максимум которой соответствовал бы величине параметра  $p_n$  (которая для всех обертонов одинакова), а ширина соответствовала бы точности реальных измерений. На рис. 1 эта функция не показана. А на всех остальных рисунках она имеется и, кроме того, пунктирной линией помечен локальный максимум результирующей функции.

Положение описанного локального максимума функции распределения практически безошибочно указывает степень роста хорошо проявившихся обертонов голоса. На рис. 2 – 8 на сонограмме виден временной курсор, и можно сравнить, как ведут себя обертона на сонограмме и максимум функции распределения параметра  $p_n$ . А на речевой паузе (на рис. 5) максимум этой функции становится существенно ниже, чем на тональных участках речевого сигнала.

Использованный в изложенном исследовании речевой сигнал имеет высокое студийное

<sup>1</sup> В действительности в эту меру включались показатели точности предсказания частоты обертона не только в соседние кадры, но и точности предсказания частоты обертона из соседних кадров в центральный.

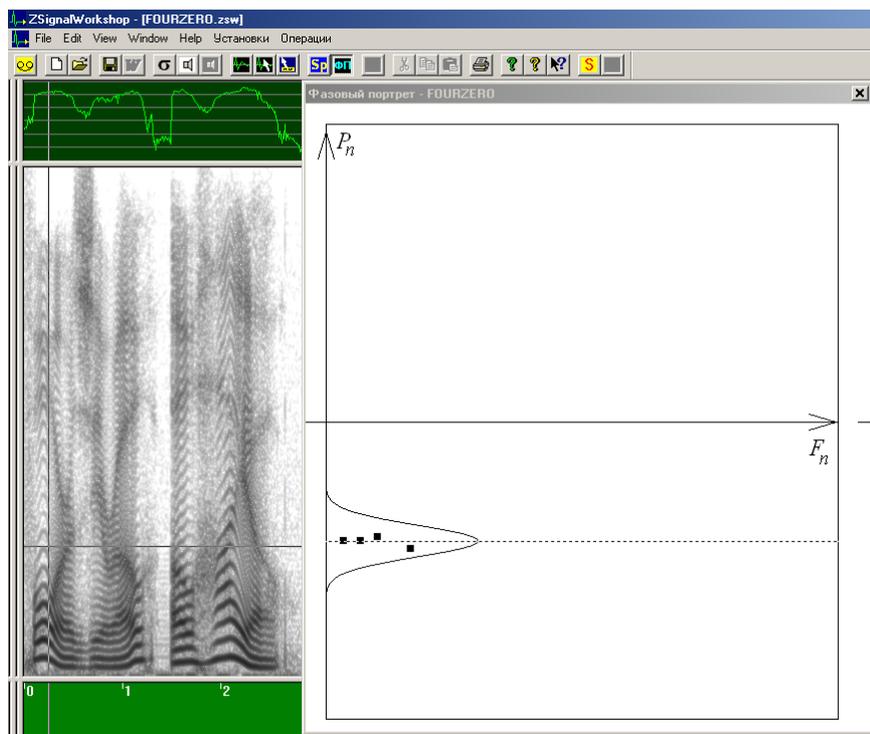


Рис. 3. Нарушение синхронности снижения частот обертонов голоса.

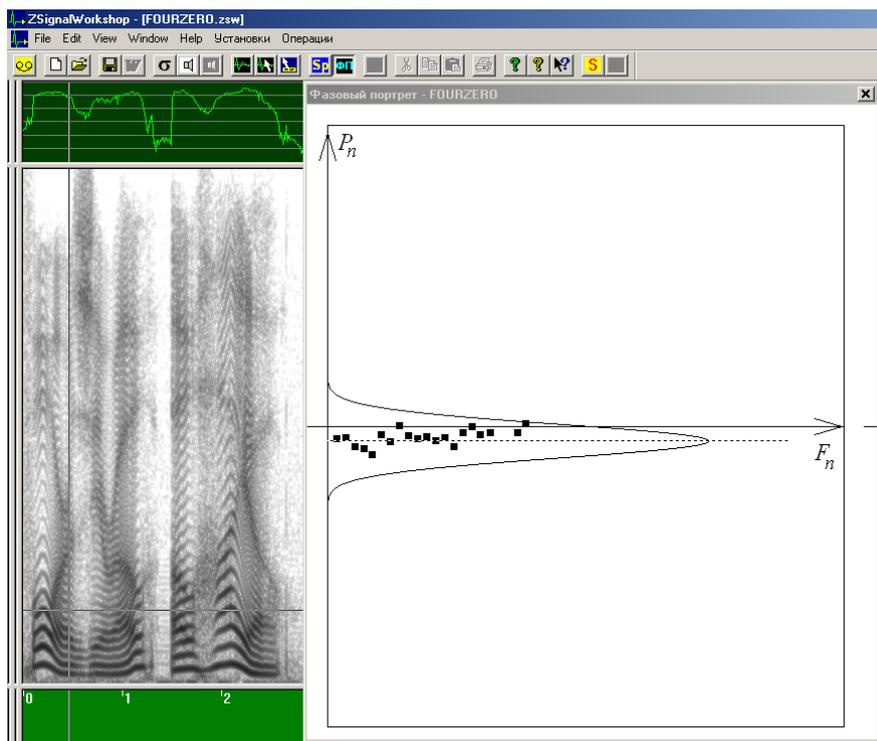


Рис. 4. Нарушение синхронности снижения частот обертонов голоса.

Описанное свойство параметра  $p_n$  может оказаться полезным не только для определения микровариаций высот обертонов голоса (например, для определения изменения эмоционального состояния говорящего). Возможно, что в динамике структуры отклонения группы обертонов от идеальной картины могут скрываться некоторые тембральные и индивидуальные особенности говорящего.

Кроме того, получаемая (выборочная) функция плотности распределения параметра  $p_n$  может играть роль дополнительного индикатора факта синхронного изменения группы гармоник по схеме обертонов голоса. А это может быть дополнительным признаком для автоматического определения динамики высоты голоса. Для этого, например, можно использовать только те значения параметра  $p_n$ , которые расположены в окрестности локального максимума полученной функции распределения. Вероятность того, что именно они соответствуют группе обертонов голоса, самая большая.

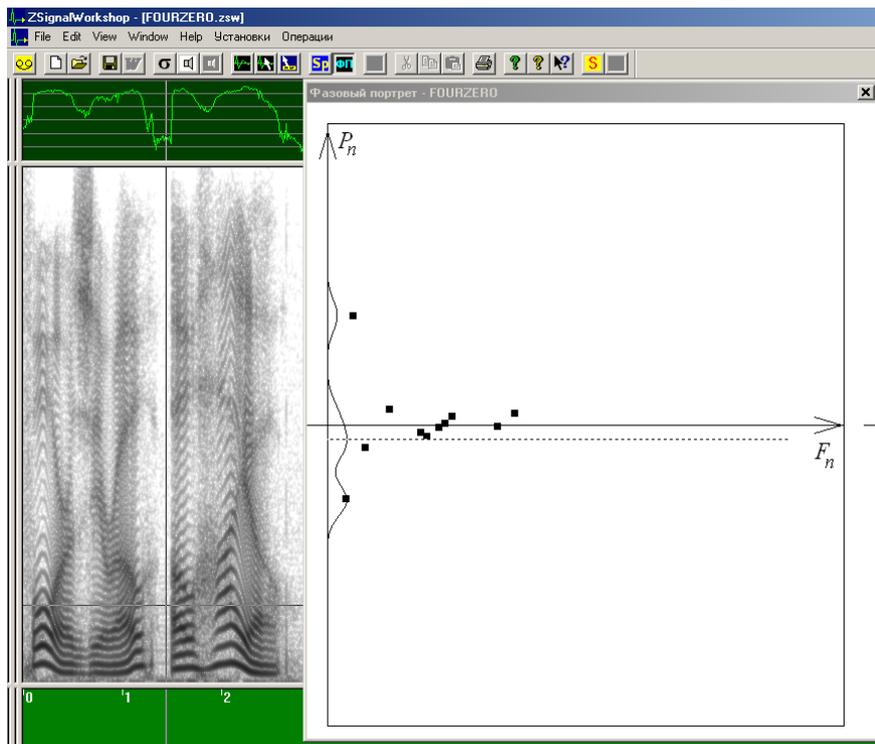


Рис. 5. Нарушение синхронности изменения частот в речевой паузе.

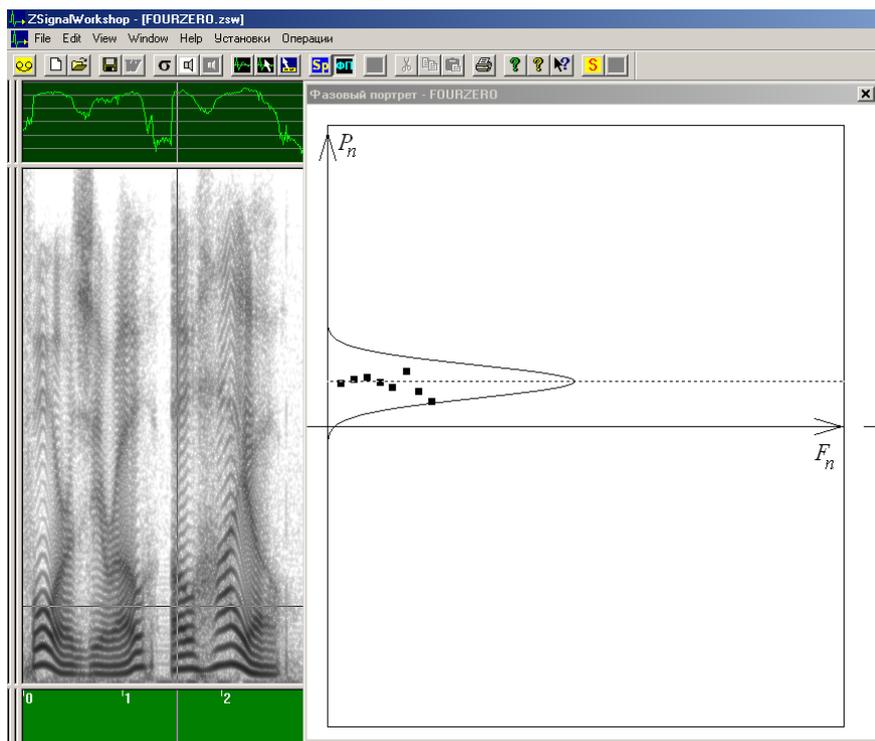


Рис. 6. Нарушение синхронности роста частот обертонов голоса.