

## Основной тон: физиология, функция в речи, измерение, интерпретация

Грамотный инструментальный анализ мелодической составляющей речевой интонации и адекватная «лингвистическая» интерпретация его результатов возможны при условии, что исследователь понимает механизм голосообразования, хотя бы в самых общих чертах, т.е. физиологические и аэродинамические процессы, происходящие в гортани во время фонации.

Может показаться, что лингвистам это совсем не нужно и не важно. Но на самом деле это и нужно, и важно, особенно тем из них, кто исследует так называемые просодические параметры речевого сигнала, потому что только хорошо представляя, как собственно образуется звук в гортани, удастся отделить изменения частоты основного тона, определяющие коммуникативно значимые компоненты интонации, от изменений, связанных с изменяющейся аэродинамической ситуацией в речевом тракте.

Это особенно важно при просодическом аннотировании звуковых файлов в речевых корпусах.

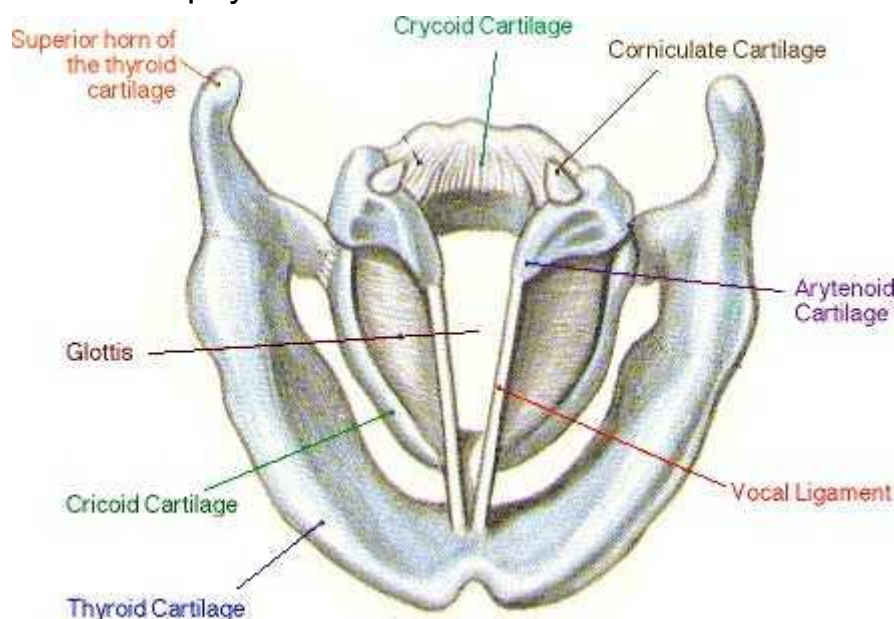


Рис. 1. Расположение основных хрящей в гортани и крепление голосовых связок

Собственно голосовые связки (vocal ligaments) представляют собой эластичные сухожилия (что-то вроде струны), одним концом прикреплённые к щитовидному хрящу (thyroid cartilage), а другим – к подвижным черпаловидным хрящам (arytenoid cartilage). К ним прикрепляются голосовые мышцы (vocalis muscle), их обволакивает

слизистая ткань, а всё вместе образует голосовые складки (vocal folds) (рис. 2).

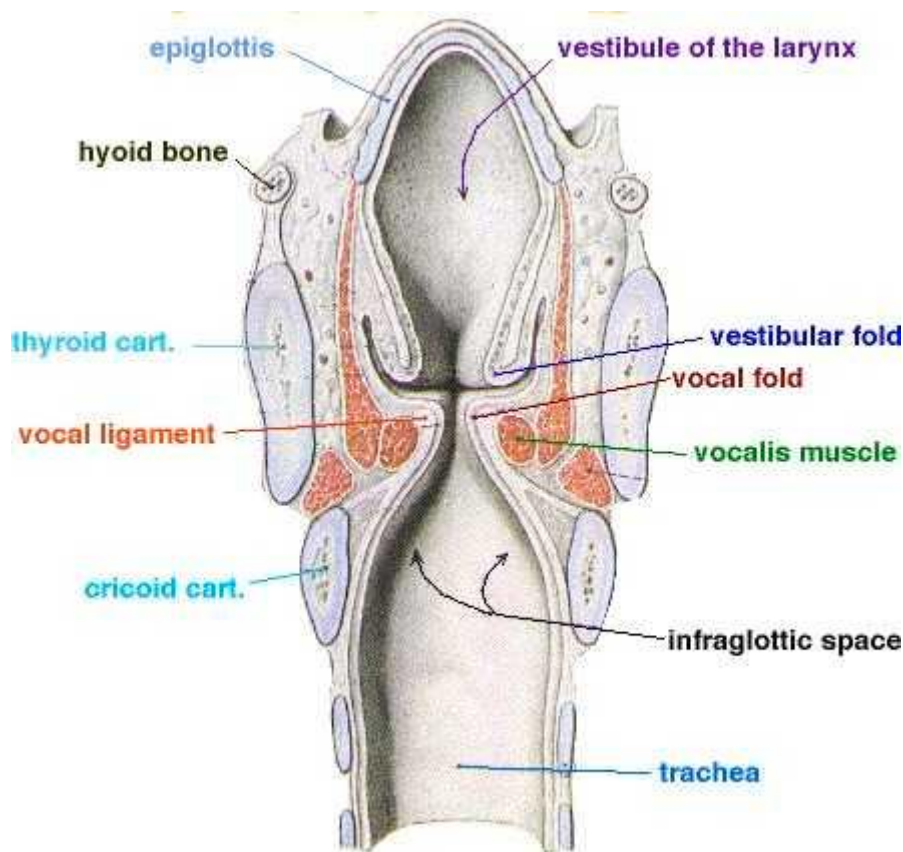


Рис. 2. Гортань в разрезе.

Чуть выше расположены ложные голосовые складки (vestibular folds), которые в определённых ситуациях исполняют некую функцию: есть мнение, что эффект гортанной смычки (glottal stop) является следствием прекращения потока воздуха в гортани из-за смыкания именно ложных голосовых складок.

Перед началом фонации, перед началом произнесения речевого сигнала сначала смыкаются черпаловидные хрящи (смыкаются голосовые складки и закрывается сзади «шепотный треугольник»), а затем начинается речевой выдох с подъёмом давления воздуха в лёгких.

За счёт того, что в лёгких давление поднимается выше, чем в окружающей среде, под напором воздуха раздвигаются голосовые складки и образуется голосовая щель (glottis), по мере развития процесса она всё больше расширяется и одновременно вступают в действие упругость самих голосовых складок и так называемый эффект Бернулли, который проявляется в том, что два предмета, между которыми идёт струя воздуха, начинают притягиваться друг к другу, и этот цикл заканчивается тем, что связки смыкаются (рис.3). Дальше

давлением воздуха в лёгких снова запускается этот процесс, т.е. возникают автоколебания в системе гортани. Длительность цикла размыкание-смыкание голосовых складок и есть тот период основного тона, исследованию которого в речи посвящено множество работ.

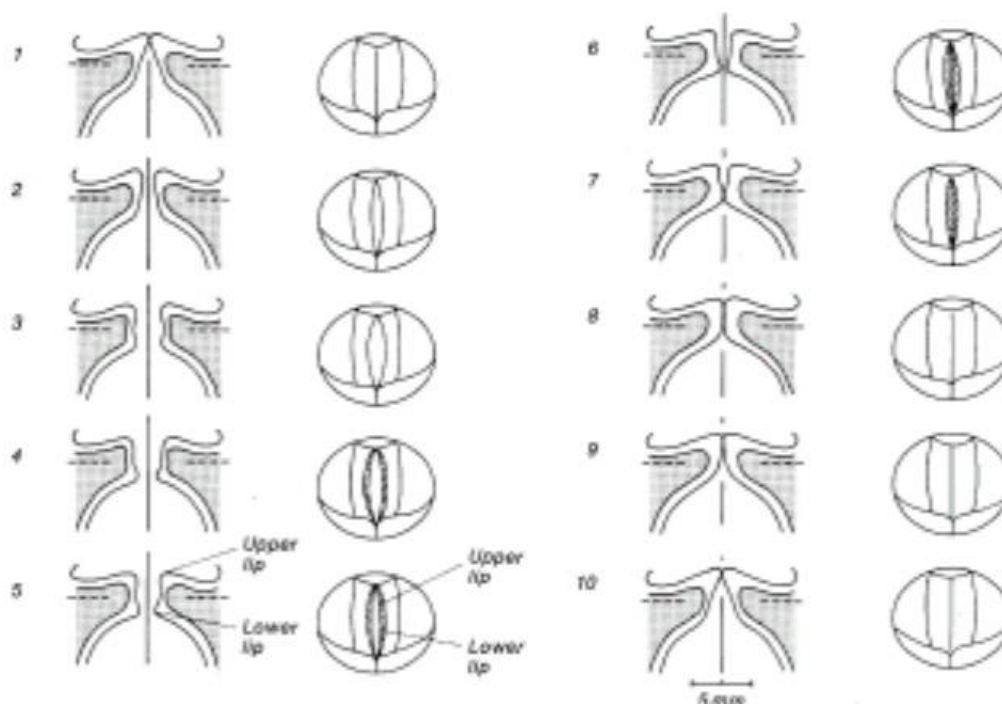


Рис. 3. Положение голосовых складок и состояние голосовой щели в течение цикла раскрытия-смыкания.

Для нормального функционирования этого процесса важным оказывается временная синхронизация процесса смыкания голосовых складок (закрытия голосовой щели) и процесса подъёма давления воздуха в лёгких. Здесь возможны, по крайней мере, три варианта.

1. Речевой выдох начинается раньше, чем закроется голосовая щель (сомкнутся голосовые связки), и вся «настройка» гортани в режим фонации происходит уже на фоне выдоха – происходит придыхательное начало.

2. Речевой выдох начинается после закрытия голосовой щели, а натяжение голосовых связок и напряжение голосовой мышцы еще недостаточны для начала фонации. Их «настройка» происходит одновременно с повышением подсвязочного давления (subglottal pressure) – происходит «мягкий» приступ.

3. Гортань полностью готова к фонации прежде, чем в лёгких образуется необходимое для этого давление, начало фонации носит взрывной характер – происходит «твёрдый» приступ. Эту ситуацию часто называют гортанной смычкой (glottal stop).

Частота образующихся автоколебаний (частота основного тона) зависит от настройки гортани: от натяжения голосовых связок, от степени

упругости голосовой мышцы, и в то же время частота колебаний напрямую связана с давлением воздуха под связками. На рис. 4 приведена зависимость частоты колебаний от величины давления воздуха в лёгких (в сантиметрах водяного столба).

В процессе речевого выдоха, т.е. при колеблющихся голосовых складках, за каждый период основного тона, когда раскрыта голосовая щель, из лёгких вытекает какое-то количество воздуха, что приводит к постепенному снижению давления в лёгких, и к понижению частоты основного тона.

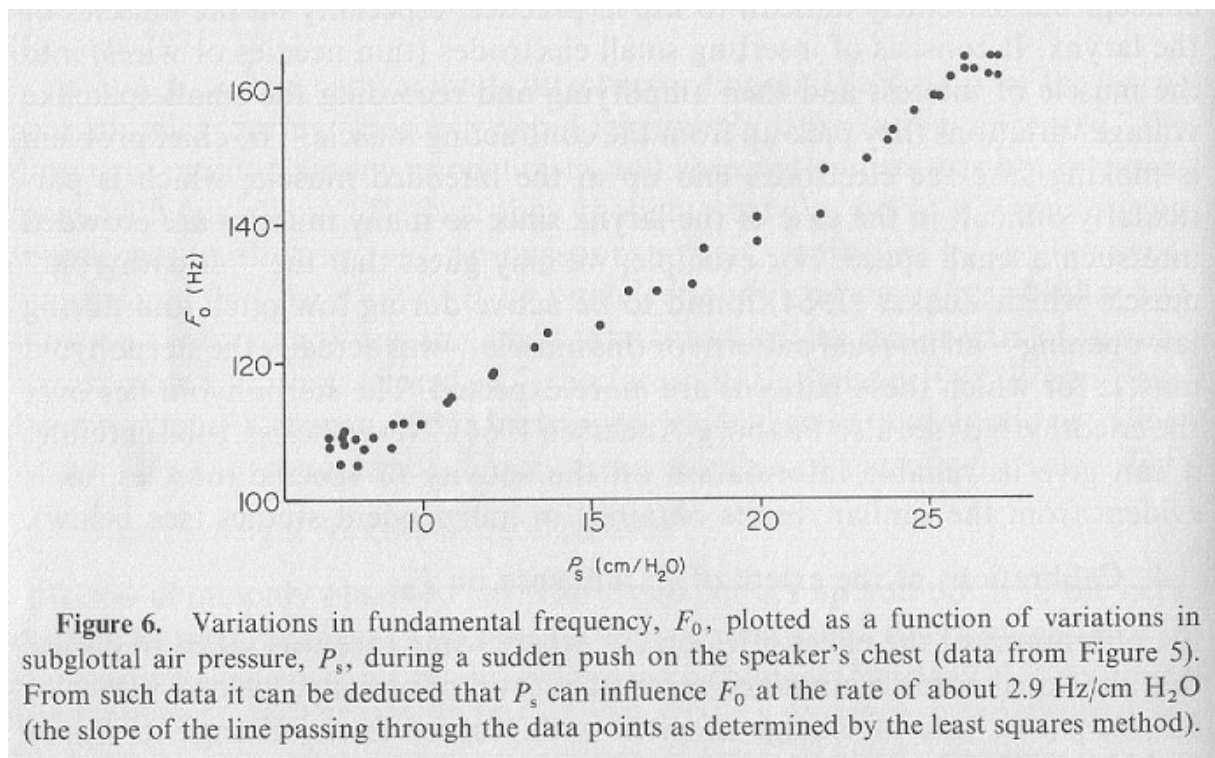


Рис. 4. Зависимость частоты основного тона от давления в лёгких (подсвязочного). По: John J. Ohala. Production of tone // Tone: a linguistic survey / V.A. Fromkin (ed.).- New York: Academic Press, 1978.

Это тот самый процесс «деклинации» частоты основного тона, исследованию которого посвящено множество работ.

Степень этого понижения давления и степень соответствующего снижения частоты основного тона зависят от интенсивности расходования запаса воздуха в лёгких.

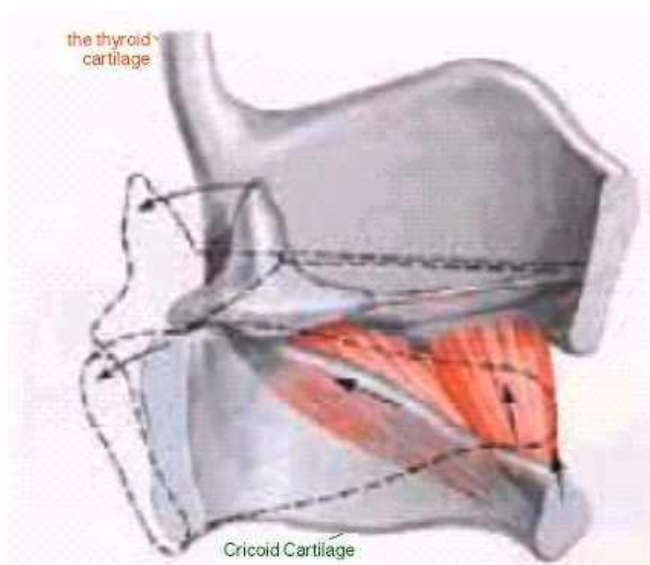
С одной стороны, интенсивность расходования воздуха может зависеть от того, насколько хорошо поставлен голос у говорящего. Как видно на рис. 3, временной интервал, в течение которого голосовая щель остаётся раскрытой, занимает только часть полного цикла колебаний. Очевидно, чем меньшую долю периода он составляет, тем экономнее расходуется воздух. Считается, что у людей с хорошо поставленным голосом, в частности у певцов, этот интервал составляет малую часть

периода. Одновременно этим обеспечивается большая интенсивность высокочастотных гармоник в спектре основного тона.

С другой стороны, в какой-то степени интенсивность расходования воздуха зависит и от сегментного состава высказывания: если в конкретном речевом сигнале много глухих щелевых согласных, при которых поток воздуха увеличивается, то, наверное, расходование воздуха и процесс деклинации будут происходить интенсивнее, чем в речевом сигнале, сплошь состоящем из гласных и сонантов, при которых потери воздуха в лёгких меньше.

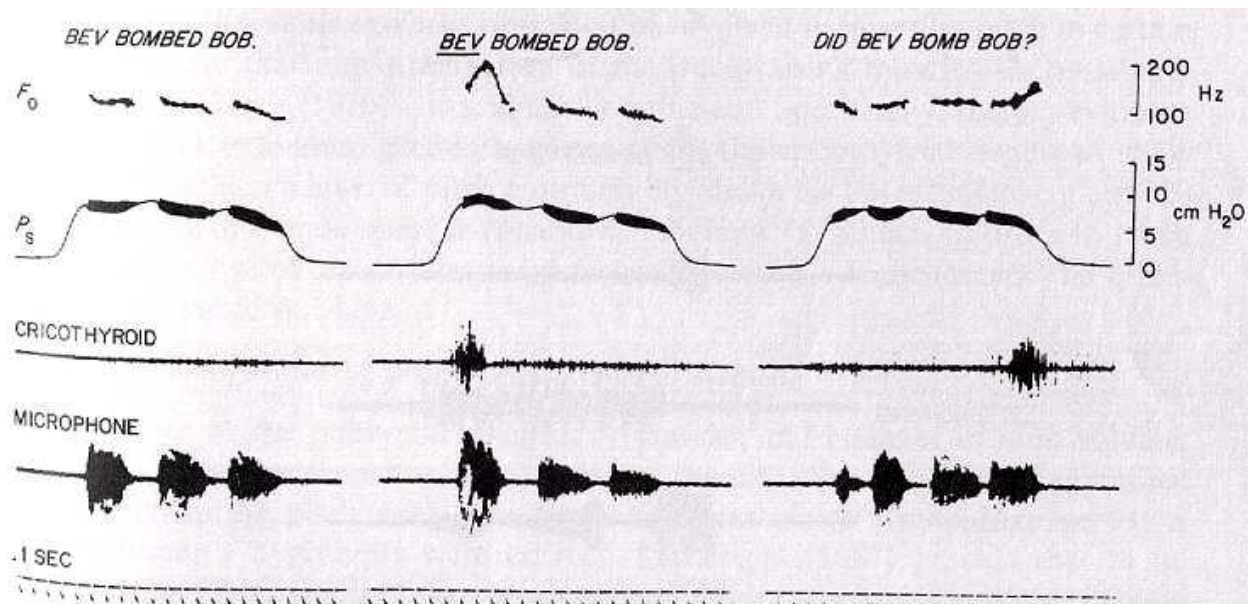
Таким образом, деклинация (понижение частоты основного тона за время звучания фразы) вероятнее всего является результатом аэродинамических процессов в речевом тракте, характеризует базовый контур частоты основного тона и не несёт никакой нагрузки в формировании коммуникативного значения высказывания.

К этому надо добавить, что в экспериментально-фонетических исследованиях вычленение этого компонента мелодического контура производится *post factum*, т.е. когда для анализа становится доступен мелодический контур всего высказывания. При этом используются математические методы обработки графиков, которые системе восприятия речи недоступны.



Сами же коммуникативно значимые мелодические изменения в речевом сигнале реализуются благодаря изменению активности группы мышц соединяющей щитовидный (thyroid cartilage) и перстневидный (cricoid cartilage) хрящи (рис. 5 и 6).

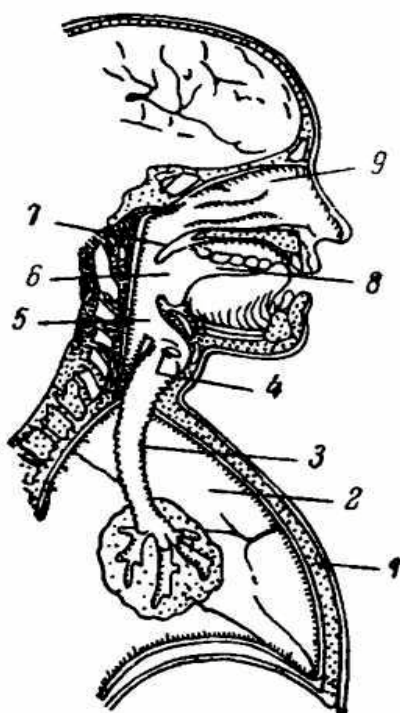
Рис. 5. Схема натяжения голосовых связок при изменении активности перстне-щитовидной мышцы.



**Figure 7.** From the top: fundamental frequency ( $F_0$ ), subglottal air pressure ( $P_s$ ), cricothyroid activity, microphone signal and .1 sec time standard during three utterances.

Рис. 6. Аэродинамическая и миозлектрическая активность в гортани при реализации высказываний с различным коммуникативным значением. По: John J. Ohala. Production of tone // Tone: a linguistic survey / V.A. Fromkin (ed.).- New York: Academic Press, 1978.

На рис. 6 хорошо видно, как синхронно с увеличением импульсации в перстне-щитовидной мышце происходит изменение частоты основного тона при неизменном подсвязочном давлении.



Всё сказанное выше относится к ситуации, когда под связками расположены лёгкие, являющиеся источником повышенного давления воздуха, а над связками – окружающая атмосфера. В реальности над связками находится речеобразующий тракт, который имеет собственный объём и содержит подвижные артикуляторы, способные создавать преграды воздушному потоку, например, при произнесении взрывных согласных. При образовании преграды начинается постепенное наполнение верхней части речевого тракта воздухом, выходящим из лёгких, разница давлений между лёгкими и ротовой полостью понижается, и соответственно понижается частота основного тона.

И вот это нужно всегда иметь в виду при

инструментальном анализе речевого сигнала: пока звучит гласный, ротовая полость открыта и влияние верхнего отдела речевого тракта на частоту основного тона минимальное; но при произнесении согласных так или иначе перекрывается поток воздуха в речевом тракте, что сразу же приводит к понижению частоты основного тона.

Результаты инструментального анализа высказывания «и более двух» (рис. 8) наглядно демонстрируют значительное понижение частоты основного тона на участках звонких согласных.

Правда, иногда в фонетических исследованиях упоминается, что характерные для большинства согласных изменения частоты основного тона могут не наблюдаться на участках носовых согласных и это позволяет считать соответствующие изменения управляемыми, т.е. коммуникативно значимыми. Однако есть исследование<sup>1</sup>, свидетельствующее, что для адекватного восприятия интонации достаточно мелодических изменений только на участках гласных.

Поскольку в инструментальных исследованиях речевого сигнала нам важны только те изменения частоты основного тона, которые несут определённую смысловую нагрузку и, следовательно, должны каким-то образом программироваться в процессе речеобразования, принимать во внимание при анализе следует только изменения частоты основного тона на участках гласных. Именно они в минимальной степени зависят от поведения артикуляторов в верхних отделах речевого тракта.

Игнорирование этого обстоятельства зачастую приводит к значительным ошибкам при просодическом аннотировании речевых корпусов (см. рис. 9) (Рассказы о сновидениях... 2008: ??).

Основываясь только на физической картине изменений частоты основного тона без учета влияния согласных была выполнена следующая просодическая аннотация: понижение с повышением с последующим понижением. Но если принять во внимание только те изменения, которые происходят на гласных, то аннотация будет совсем другая: подъём на «полетели» и понижение на «вверх» либо высказывание следует считать двухсинтагменным, исходя из чрезмерной длительности согласного «в». И, к сожалению, в тех корпусах, где есть просодическая аннотация, такие несуразности не редки.

---

<sup>1</sup> Венцов А. В., Нушикян Э. А. Исследование возможности восприятия мелодического контура, представленного изменениями основного тона только на гласных // Матер. докл. и сообщений 5-го Всесоюзного совещания-симпозиума цикла «Акустика речи и слуха: Эмоции и автоматическое распознавание речи». Одесса, 1989.

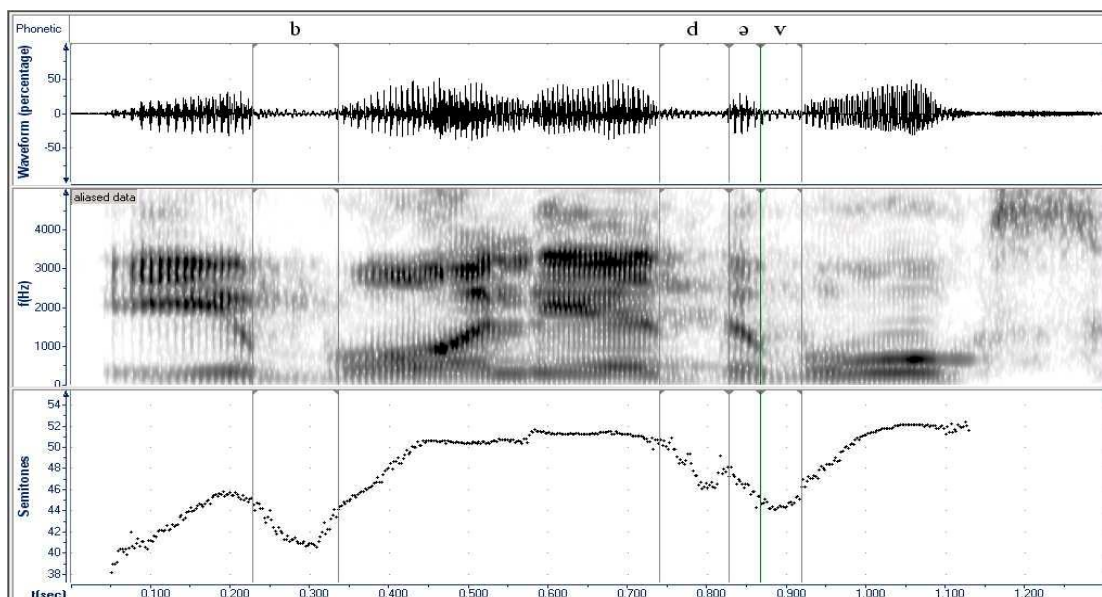


Рис. 8. Осциллограмма, динамическая спектрограмма и мелодический контур высказывания *и более двух*.

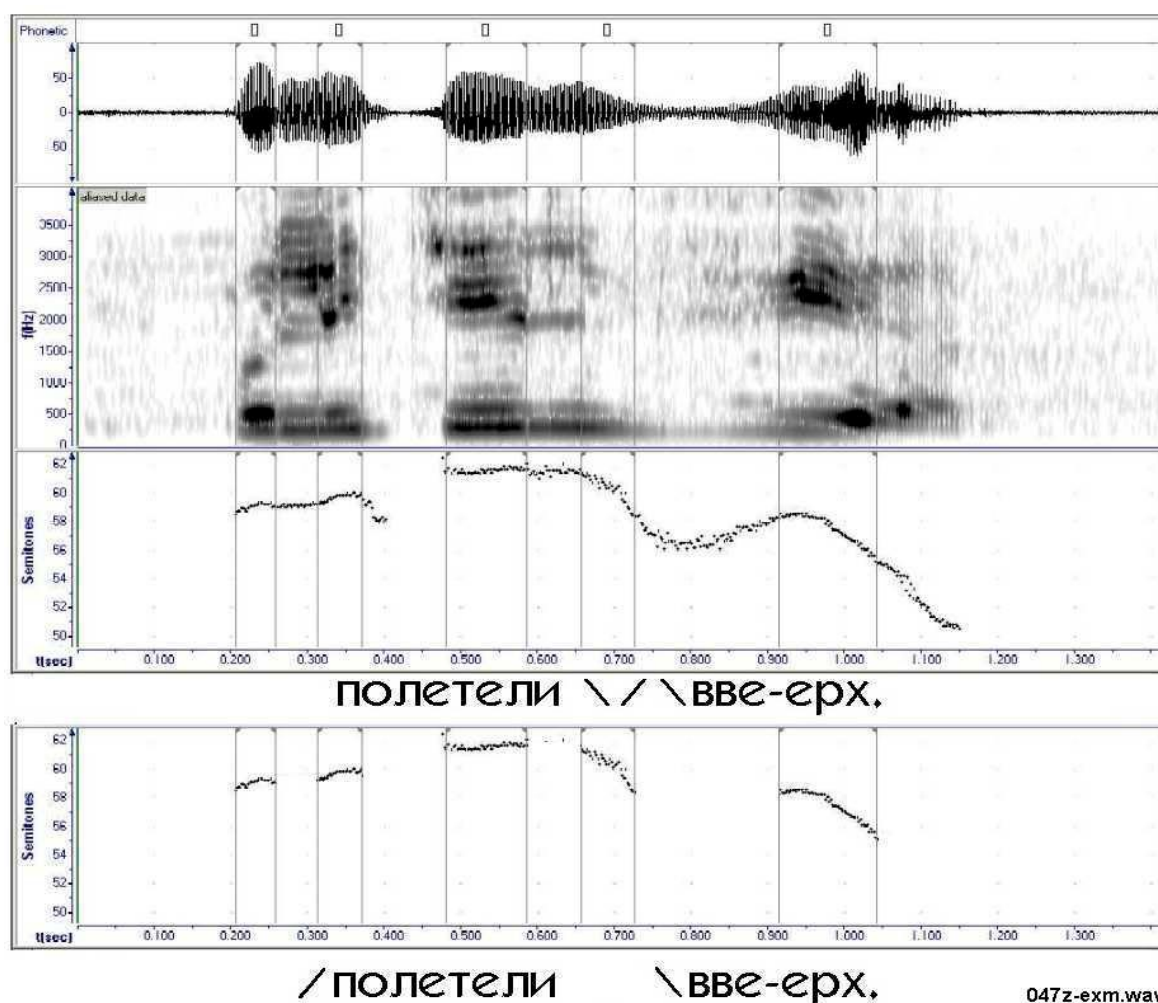


Рис. 9. Результаты инструментального анализа фразы «полетели вверх» (Рассказы о сновидениях) и примеры просодической аннотации.

Какие бы участки мелодического контура ни принимать во внимание при «лингвистической» интерпретации результатов инструментального анализа, эти результаты надо прежде всего как-то получить, т.е. надо из исследуемого речевого сигнала выделить основной тон, точнее, получить график изменения его частоты во времени.

Чаще всего для этой цели используют компьютерные программы SpeechAnalyzer и PRAAT.

Обе программы используют алгоритмы выделения основного тона, использующие такое свойство речевого сигнала (колебаний голосовых складок) как периодичность, точнее, квази-периодичность, поскольку даже на тех участках сигнала, где исследователи фиксируют «ровный тон», два соседних периода колебаний зачастую оказываются разной длительности, но в чём-то очень «похожими». Алгоритмически степень «похожести» может быть оценена коэффициентом корреляции двух соседних выбранных участков речевого сигнала: чем выше этот коэффициент, тем более «похожими» можно считать сигналы на этих участках.

Как это получается на практике? В сигнале выделяется некоторый кусочек, отрезок какой-то длительности («окно»). Его «ширина» (длительность) выбирается исследователем (или жестко задана в конкретной компьютерной программе) и зависит от того, с какой подробностью желательно получить график изменения частоты основного тона: чем больше (шире) будет это окно, тем более сглаженной будет результирующая кривая.

Затем вычисляется коэффициент корреляции выбранного отрезка сигнала с частью сигнала равной длительности, но сдвинутого относительно начала исходного отрезка. Операция осуществляется шаг за шагом с последовательным увеличением длительности сдвига и останавливается при достижении максимума коэффициента корреляции. Длительность соответствующего сдвига принимается в качестве оценки периода основного тона.

Дальше «окно» сдвигается на величину временного шага, с которым исследователь предполагает получать оценки частоты основного тона, и процедура вычисления коэффициента корреляции (и оценки периода основного тона) повторяется. В результате исследователь получает график изменений частоты основного тона с шагом ШШ при «окне» ОО. Следует отметить, что в программе SpeechAnalyzer и «окно», и шаг жестко заданы, и об их конкретной величине нет информации даже в описании программы. В программе PRAAT «ширина окна» равна шестикратной длительности шага, при этом последнюю можно изменять, но каким образом это сделать, понять совсем не просто.

Особенность описанного алгоритма является вероятность того, что при каких-то параметрах анализируемого сигнала и «настройке» программы анализа максимальная величина коэффициента корреляции моет

достигаться либо при половинной длительности ожидаемого периода основного тона, либо при двойной. В обоих случаях на графике изменений частоты основного тона возникает так называемый «октавный скачок».

В программных пакетах PRAAT предлагается инструмент (скрипт) для нейтрализации такого «октавного скачка».

Но важно понимать, что октавный скачок может быть как следствием особенностей алгоритма, с помощью которого выделяется основной тон из речевого сигнала (измеряется частота О.Т.), так и результатом изменений в функционировании голосовых складок на данном отрезке речевого сигнала (ларингализация). В последнем случае принудительная нейтрализация октавного скачка скроет информацию об этом. И очень важно иметь инструмент, который позволяет оценить, является ли октавный скачок ошибкой «прибора» (алгоритма) или у говорящего на самом деле произошло временное понижение (частоты) основного тона на октаву (см. рис. 10 вверху).

Здесь стоит вспомнить, что основной тон – это первая гармоника сложного гармонического сигнала, а его частота – частота первой гармоники или разность частот двух соседних гармоник. Это даёт возможность проверять сомнительные случаи, используя для анализа сигнала спектральный анализатор, способный выделять гармоники.

В программе SpeechAnalyzer предусмотрена возможность изменять параметры спектрального анализатора: вместо широкополосных фильтров использовать узкополосные. Можно также изменить диапазон частот, представляемых на графике (см. рис. 10 внизу). И теперь мы видим, что на спектрограмме проявляются не форманты, а гармоники сигнала, и что на участке октавного скачка гармоники идут вдвое чаще, чем на предшествующем сегменте сигнала, т.е. на самом деле мы имеем дело не с октавным скачком, вызванным особенностями алгоритма, а с понижением тона у говорящего на последнем слове фразы.

Подобный алгоритм анализа основного тона реализован в программе WinSnoori: в ней частота основного тона измеряется по расстоянию между гармониками.

Таким образом, оказывается, что для получения надёжных результатов инструментального анализа речевого сигнала исследователю желательно иметь в своём распоряжении несколько программ анализа, чтобы в сомнительных случаях сопоставлять результаты работы разных алгоритмов.

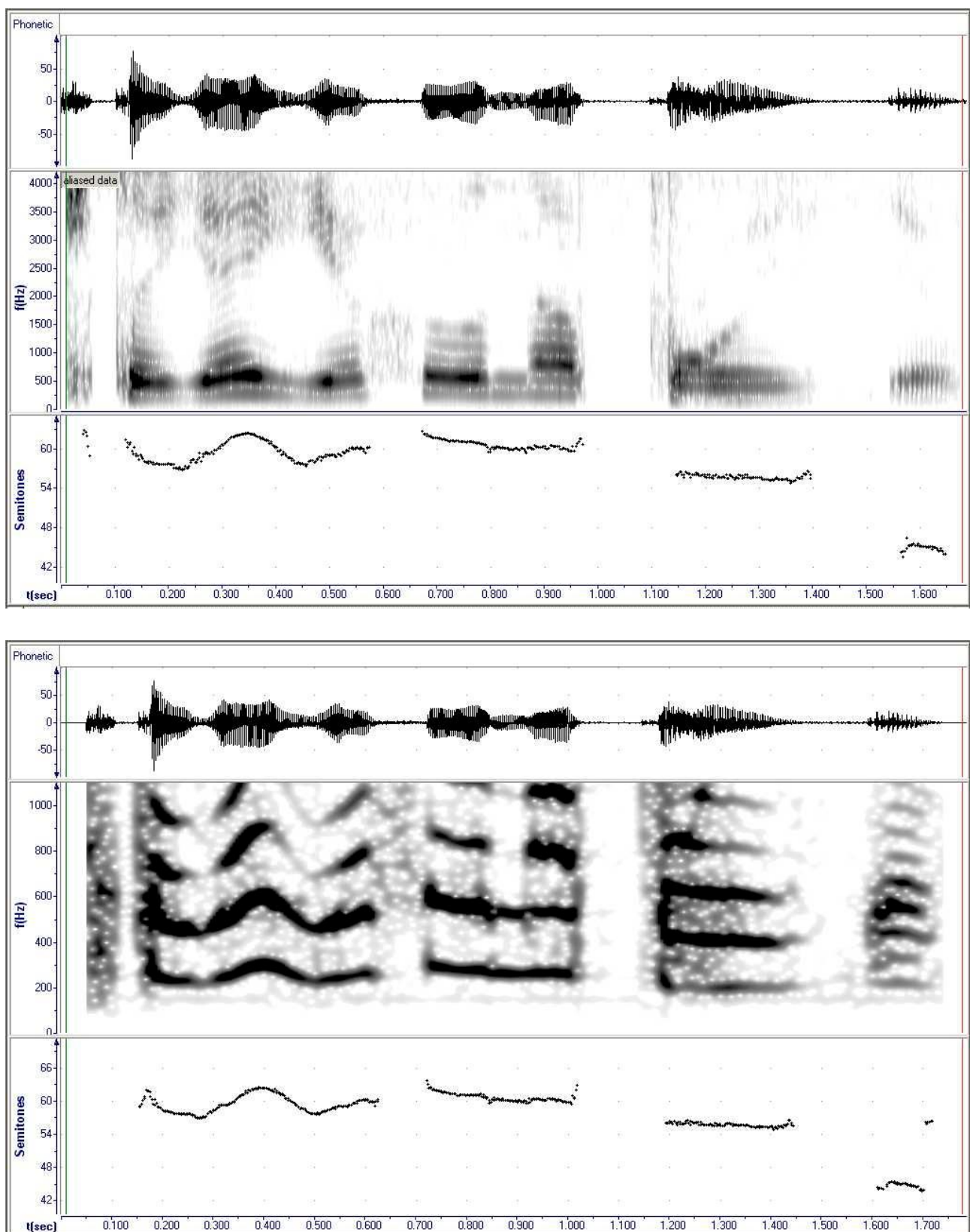


Рис. 10. Результаты инструментального анализа фразы «что я ехала на поезде» (048z-ext.wav [Рассказы о сновидениях...]). Вверху – широкополосный анализатор, внизу – узкополосный.

Какова же роль изменений частоты основного тона в системе языков? Судя по тому, что есть в литературе, изменения частоты основного тона задают коммуникативное значение высказывания или лексическое значение единицы языка (в тональных языках), а также эмоциональную окраску дискурса.

Применительно к русскому языку система интонационных единиц (контуров) и связанные с ними коммуникативные типы высказываний детально описаны в работах Е.А. Брызгуновой.

И вот тут возникает вопрос: что считать изменением частоты основного тона и с какой точностью следует измерять эту частоту?

Использование вычислительной техники при инструментальном анализе речевого сигнала создаёт обманчивое впечатление высокой точности получаемых результатов. К примеру, PRAAT выдаёт частоту основного тона с множеством знаков после запятой. Как получаются такие числа?

Предположим, что при определении периода основного тона корреляционным методом получилась длительность в 200 шагов дискретизации (именно такими шагами происходит сдвига «окна», о котором речь шла выше).

Длительность шага дискретизации при частоте 22050 Гц компьютер вычислит как 0,045351473922902494331065759637188 мс. Тогда период основного составит 9,0702947845804988662131519274376 мс.

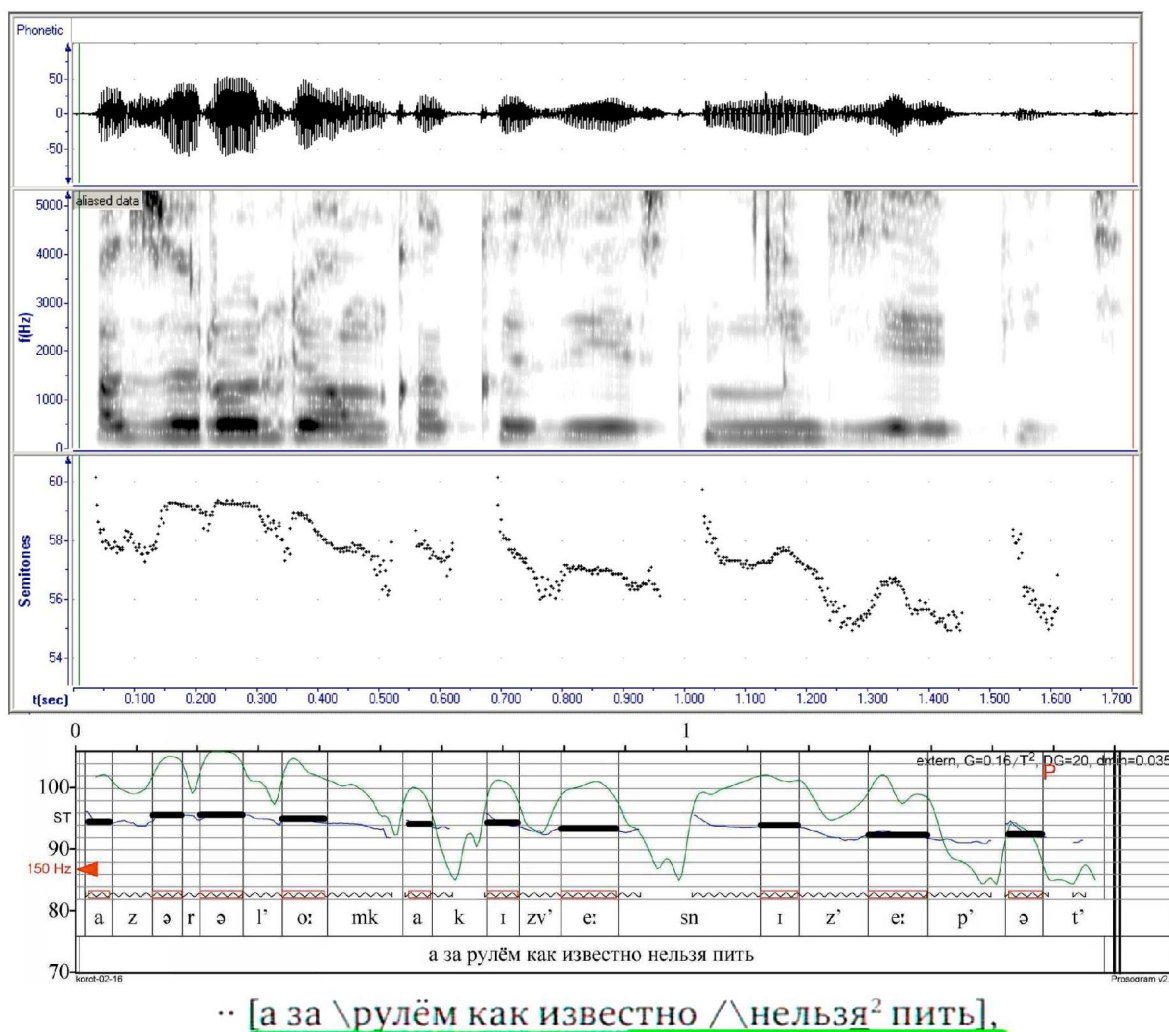
Компьютерный калькулятор работает с округлением и выдаёт при такой длительности периода частоту 110,25 Гц. Если в программе PRAAT округление не предусмотрено, количество знаков после запятой окажется огромным.

Конечно, фонетисты при описании частоты основного тона в герцах вольны как угодно округлять полученные в результате такого измерения величины и их изменения во времени, поскольку они описывают систему порождения речи и такие данные могут затем использоваться в процедурах синтеза речи по тексту, когда желательно получить максимально естественно звучащую синтезированную речь. Но если речь идёт о системе восприятия речи, приходится принимать во внимание свойства слуховой системы человека, и тогда оказывается, что обнаруженные инструментальными методами изменения могут не быть таковыми для слушателя. Это относится и к ступенчатым (регистровым) изменениям частоты основного тона, и к плавным (контурным).

Психоакустические данные свидетельствуют, что ступенчатые изменения частоты основного тона на два полутона на самом деле слуховой системой человека не воспринимаются.

Не проще обстоит дело с плавными изменениями частоты.

Экспериментальные исследования показали, что при некоторых параметрах этих изменений слуховая система их не обнаруживает.



Pic-RUS\_01-f\_Ski-T-16

Рис. 11. Сопоставление результатов инструментального анализа фразы «а за рулём, как известно, нельзя пить»<sup>2</sup> и предполагаемого отражения их в слуховой системе человека.

Наверху – результаты инструментального анализа; внизу – их представление в программе PROSOGRAM.

Свойства слуховой системы при обработке основного тона учтены в программном пакете PROSOGRAM написанном Питом Мертенсом для PRAAT'а, которым выполняется некоторая аппроксимация (стилизация) реального мелодического контура так, как если бы это делала слуховая система.

<sup>2</sup> Кортаев Н. А. Коммуникативно-просодический подход к выявлению элементарных дискурсивных единиц в устном монологическом тексте // Компьютерная лингвистика и интеллектуальные технологии. По материалам ежегодной Международной конференции “Диалог” (2015). Выпуск 14.– С. 306-320.

Как следует из рис. 11, плавные (контурные) изменения частоты основного тона, видимые в физической картине, отсутствуют в «слуховом» представлении, а ступенчатые – не превышают двух полутонов. Приходится сделать вывод, что данное высказывание произнесено совершенно монотонным голосом, хотя в работе, откуда взята ссылка на соответствующий акустический сигнал, просодическая аннотация изобилует движениями тона.

Поэтому в исследованиях восприятия носителями языка мелодических изменений в речи, а для лингвистов как раз это самое важное, предпочтительнее пользоваться именно таким представлением.

На рис. 12 приведены результаты обработки фразы «это изменит ситуацию», которая аудиторам воспринимается, как вопросительная.

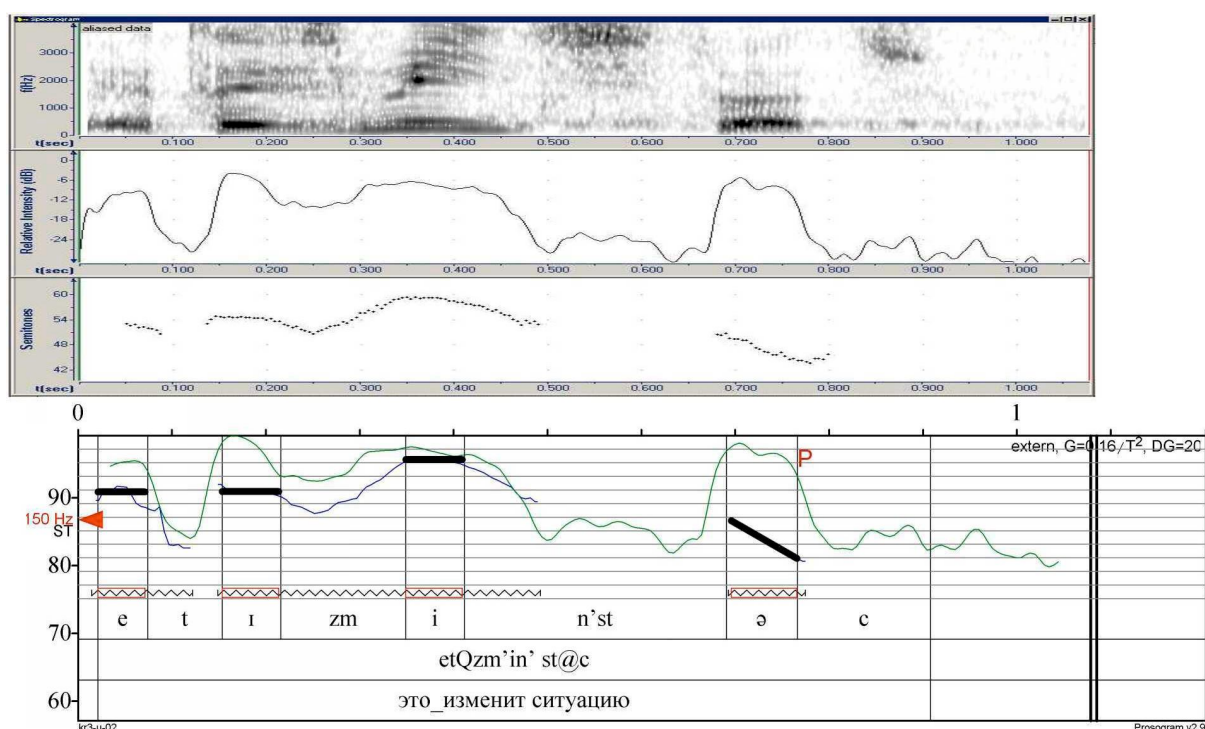


Рис. 12. Результаты анализа фразы «это изменит ситуацию» программами SpeechAnalyzer (вверху) и Prosogram (внизу).

Судя по приведенным графикам, можно сказать, что для восприятия вопроса достаточно повышения частоты основного тона на ударном гласном на четыре полутона с последующим понижением на несколько полутонов.

Но если повышение частоты основного тона на ударном гласном последнего слова во фразе не сопровождается последующим понижением (высокий тон сохраняется до конца), фраза по Е.А. Брызгуновой должна восприниматься, как незавершенная (рис. 13).

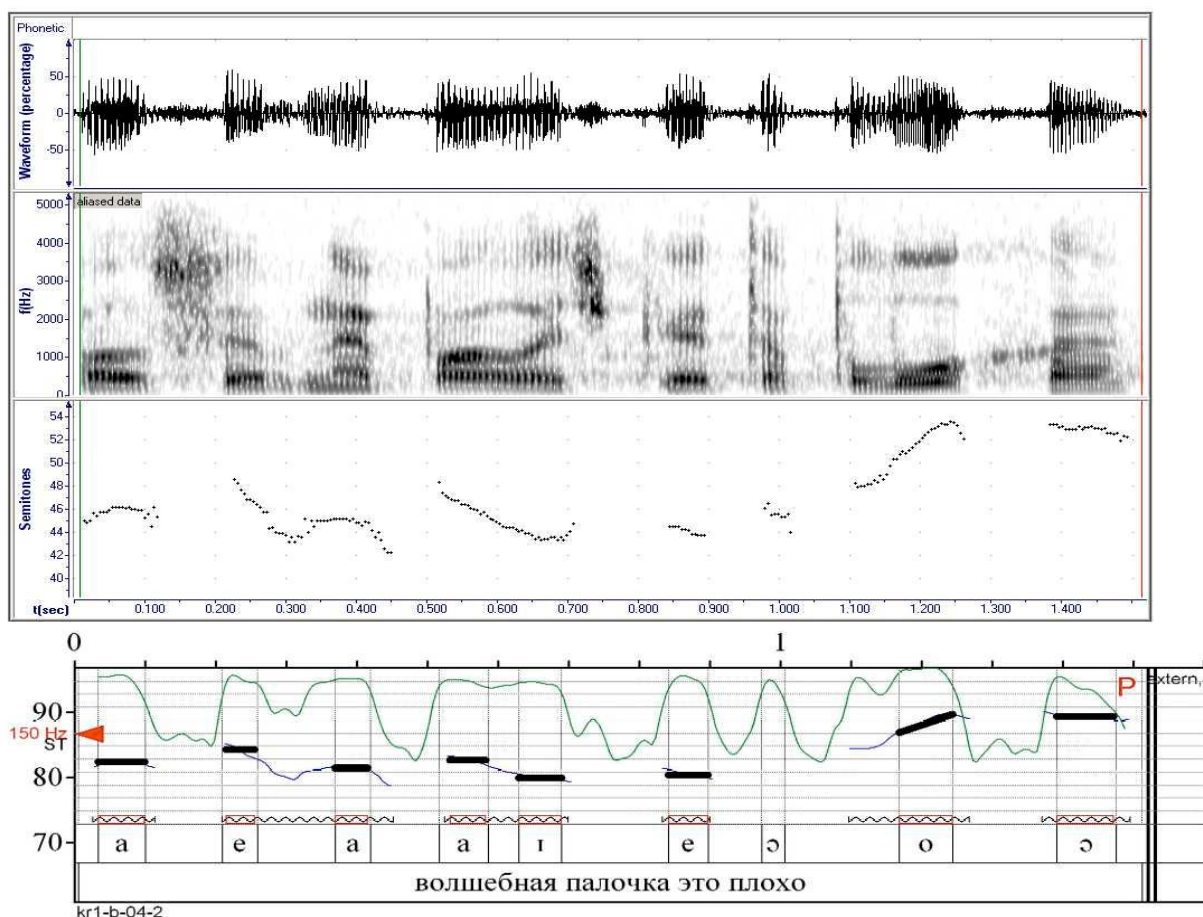


Рис. 13. Результаты анализа фразы «волшебная палочка это плохо» программами SpeechAnalyzer (вверху) и Prosogram (внизу).

Надо ещё сказать, что сама конкретная интерпретация мелодического контура зависит не только от собственно мелодических изменений, которые мы измерим в речевом сигнале, но и от лексического, лексико-грамматического состава высказывания, на которое это изменение накладывается.

Так, мелодический контур на рис. 14 может быть интерпретирован как проявление интонации вопроса или незавершенности. Однако в составе целостной синтагмы он очевидным образом реализует акцентную выделенность глагола «пригласили» (рис. 15). Таким образом, один и тот же контур в зависимости от окружения может восприниматься либо как вопрос, либо как логическое выделение.

Предполагается, что помимо реализации чисто коммуникативных значений интонация (её мелодический компонент) может выполнять делимитативную функцию, т.е. быть источником пограничных сигналов между следующими друг за другом синтагмами, чем обеспечивается надёжность их восприятия слушателем. В качестве такового предлагается «ресет» (перезагрузка), или «скачок» частоты основного тона в линии деклинации на месте предполагаемой границы (рис. 16).

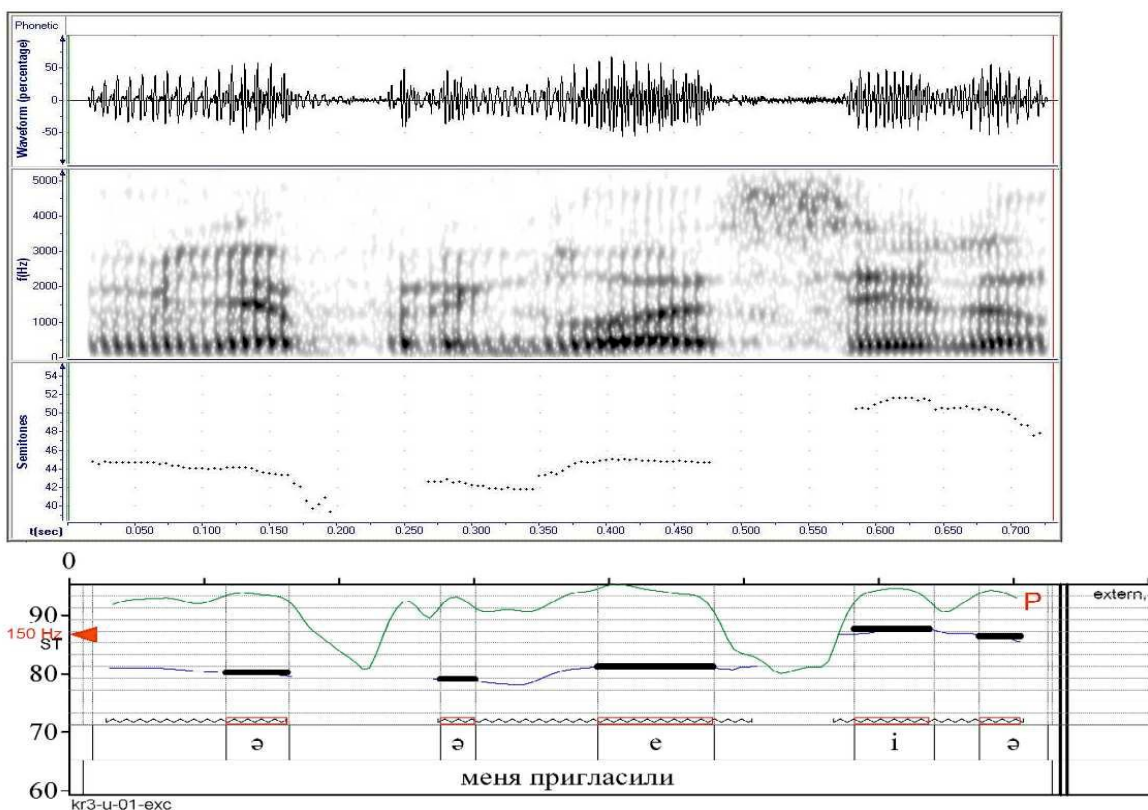


Рис. 14. Результаты инструментального анализа фразы «меня пригласили». Сверху вниз: SpeechAnalyzer и Prosogram.

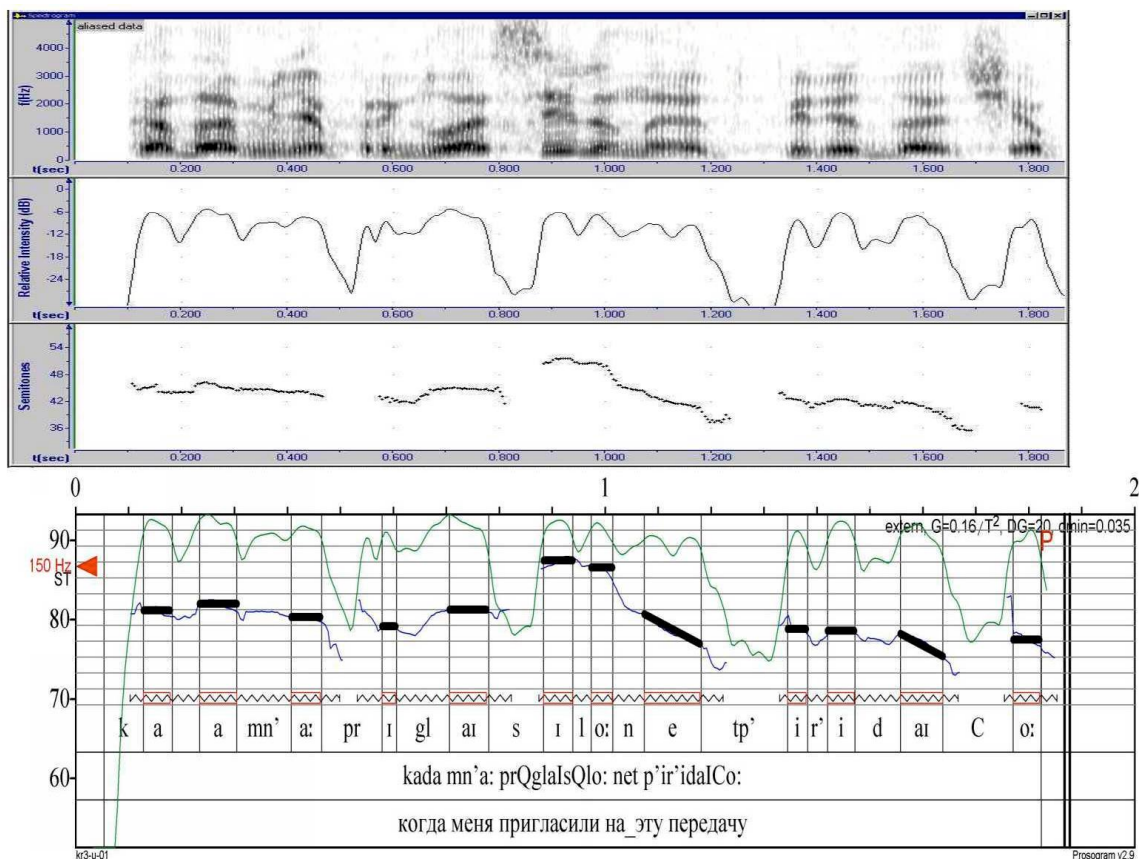
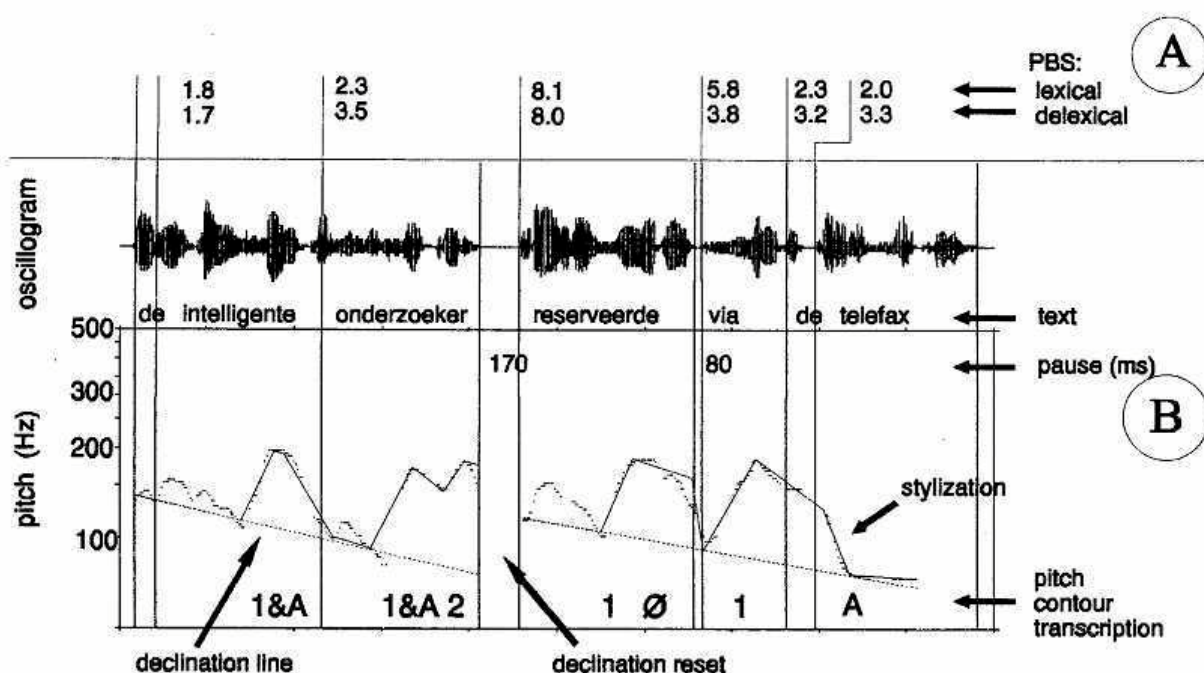


Рис. 15. Результаты инструментального анализа фразы «когда меня пригласили на эту передачу». Сверху вниз: SpeechAnalyzer и Prosogram.



**Figure 1:** (A) PBS values and (B) results of the phonetic analyses for one of the test utterances of the professional speaker.

Рис. 16. Иллюстрация «делимитативной» функции интонации.

На приведенном рисунке мелодические контуры представлены двумя составляющими: некоторым «базовым» изменением частоты основного тона (линия деклинации - понижение от начала к концу фразы) и наложенными на него изменениями, которые связаны с коммуникативными и смысловыми характеристиками конкретного высказывания.

Чтобы выделить эти составляющие исследователь должен анализировать отрезки речевого сигнала значительной длительности и при этом применять некоторые «математические» методы.

Слуховая система носителя языка лишена возможности оперировать одновременно столь длительными порциями речевого сигнала и не имеет в своём арсенале инструментов для подобного «математического» анализа.

Таким образом, раздельный анализ составляющих мелодического контура с фиксацией резких изменений в параметрах линии деклинации представляется вполне обоснованным при исследовании процессов речеобразования (с возможностью последующего использования при синтезе речи по тексту), но нет никаких оснований полагать, что этими же параметрами пользуется система восприятия речи.

К тому же, следует отметить, что «ресет» на рис. 16 сопровождается паузой в речевом сигнале. Очевидно, завершив одну синтагму, говорящий начинает следующую как бы с самого начала, и весь процесс подготовки артикуляторного тракта и лёгких начинается заново.

К сожалению, в подавляющем большинстве работ, исследующих «ресет» линии деklinации, не упоминается наличие/отсутствие паузы в месте ожидаемой границы, маркером которой предположительно и является «ресет». В то же время при анализе спонтанной речи обнаруживается множество случаев, когда даже пауза на стыке двух дискурсивных единиц не сопровождается существенными изменениями частоты основного тона (рис. 17).

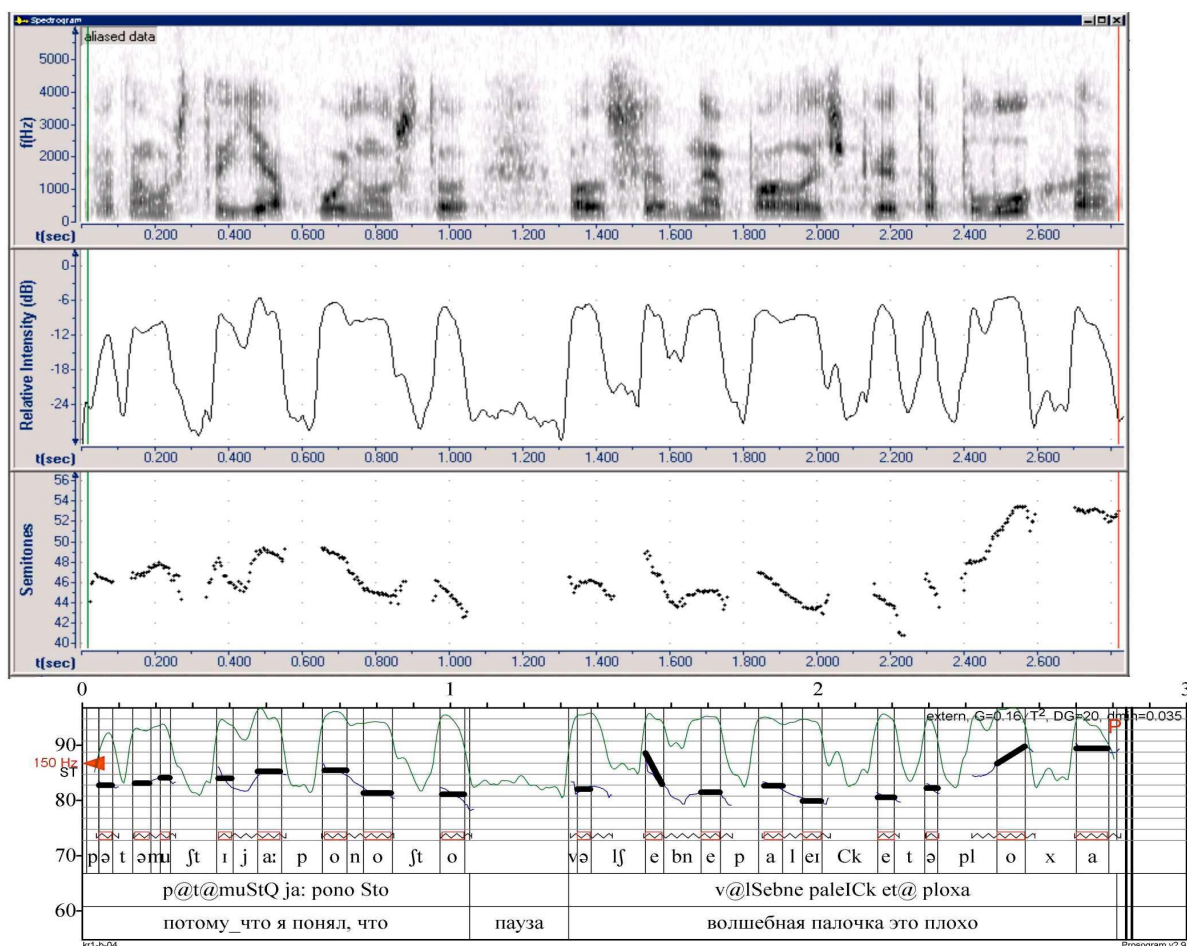


Рис. 17.

Но встречаются случаи, когда две дискурсивные единицы произносятся слитно, и тогда на их стыке не должно быть «ресета» теоретически, нет его и практически (рис. 18), поскольку обе единицы произносятся в едином цикле речевого выдоха.

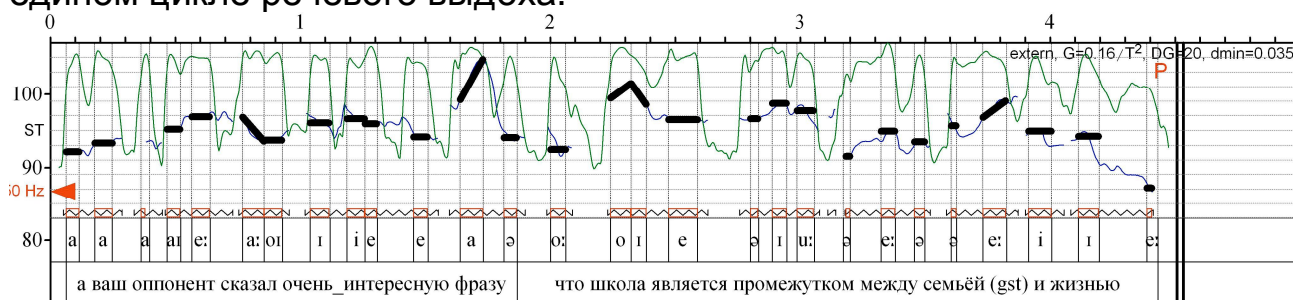


Рис. 18. Просограмма высказывания «а ваш оппонент сказал очень интересную фразу, что школа является промежутком между семьёй и жизнью».

Наконец, возможны ситуации, когда дискурсивная единица произносится в виде нескольких изолированных высказываний и каждое имеет четкую индивидуальную интонационную оформленность (рис. 19), так что при прослушивании каждое из них может быть воспринято самостоятельной дискурсивной единицей (в определенном контексте, разумеется).

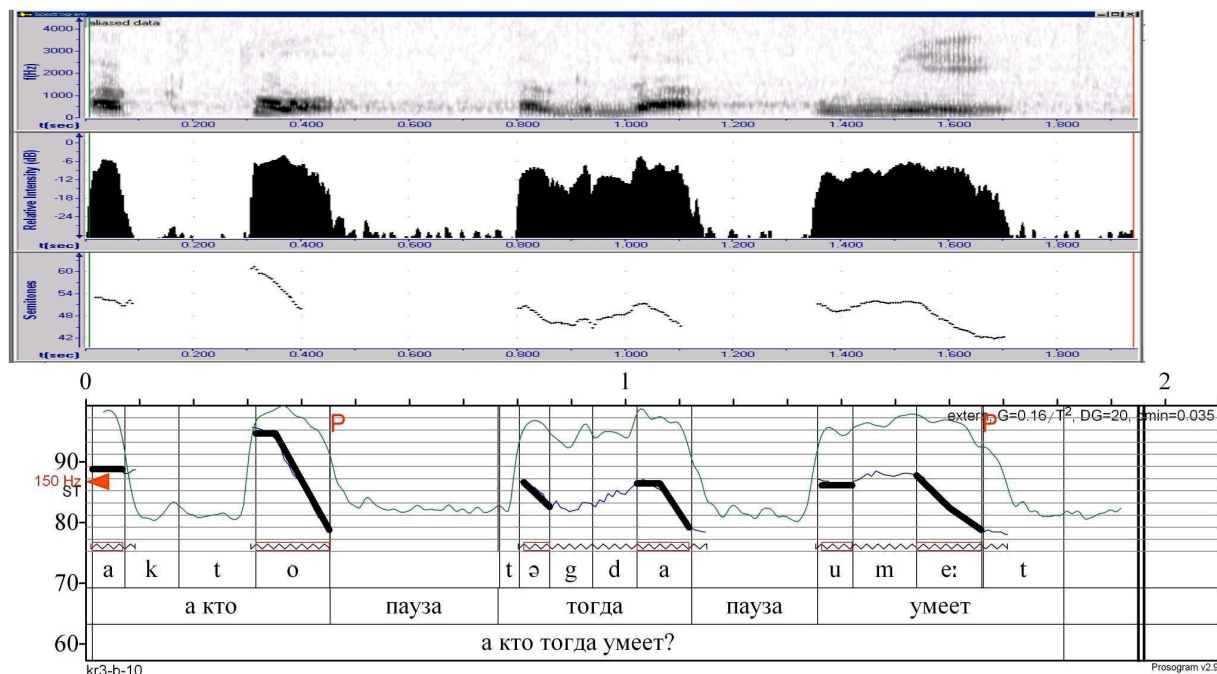


Рис. 19.

Не следует ли из всего этого, что в естественном речевом сигнале может не быть границ между дискурсивными единицами (клаузами), маркированных мелодическим компонентом интонации?

При этом создатели систем автоматической постановки границ при аннотировании речевого сигнала обнаруживают, что самым важным признаком является наличие паузы [Mertens et al. 2013].

Но пауза, как показано выше, может членить клаузу на более мелкие отрезки, хотя во всех подобных случаях носители языка совершенно адекватно воспринимают речь говорящего.

Возникает задача исследовать правила, пользуясь которыми при восприятии речи носитель языка разделяет слитно произнесенные клауза и объединяет части «раздробленных» клауз, и поскольку выделение фонетических маркеров границ проблемы не решает, следует предположить, что слушатель опирается на имеющиеся у него знания о языке и текущей коммуникативной ситуации, обходясь без фонетической информации.

И всё же фонетисты стремятся маркировать границы, опираясь на характерные особенности мелодического контура. Подчас это приводит к курьёзным результатам (рис. 20).

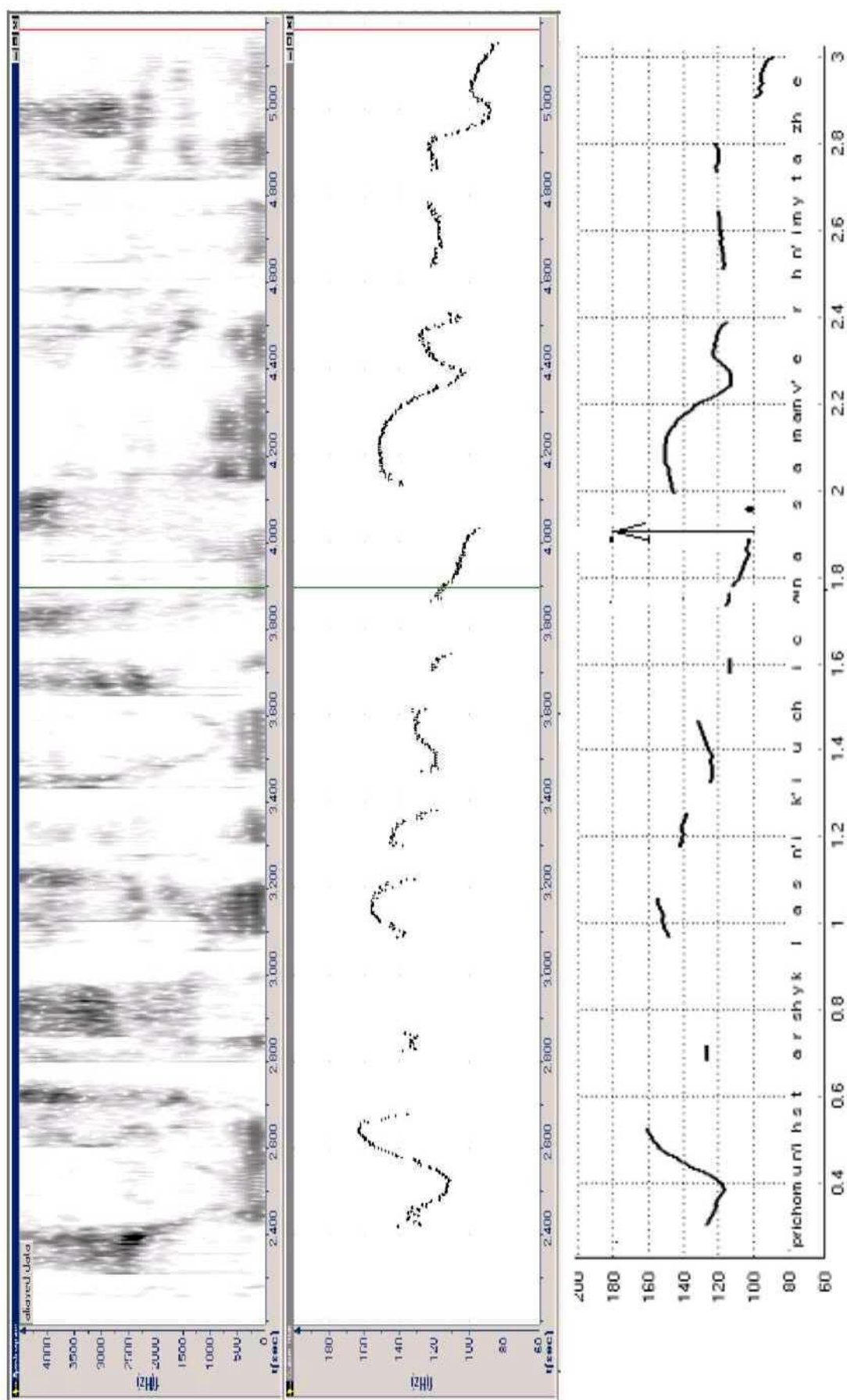


Рис. 20. Результаты анализа фразы «причём у них старшеклассники учатся на CAMOM верхнем этаже» (av18t-1.wav).

Так в одной из работ момент резкого изменения частоты основного тона (отмечен стрелкой) был принят за тот самый «ресет», хотя на самом деле никакой границы между синтагмами здесь нет. «Скачок» основного тона, видимый только в соответствующем контуре, образовался из-за того, что тонально подчеркнутому гласному выделенного слова «самом» предшествует глухой щелевой согласный.

Теперь ещё несколько слов по поводу так называемых тональных языков.

«В говоре дер. Ваньцзы шесть тонов: 1) высокий ровный (тон 55); средне-высокий ровный (тон 44); средний ровный (тон 33); средний восходящий (тон 24); низкий восходящий (тон 13); низкий нисходящий (тон 31). Тон 33 обычно произносится не совсем ровно, а с небольшим понижением контура (43)» [Самарина и др. 2011: 47]<sup>3</sup>.

Из этого следует, что в этом говоре существуют три регистровых тона, т.е. при одинаковом сегментном составе слога его лексическое значение определяется высотой реализованного тона.

Неизбежно возникает вопрос: как должны реализовываться и как могут восприниматься такие слоги в произнесении мужчин и женщин?

Дело в том, что женские голоса, как правило, выше мужских, т.е. частота основного тона у женщин выше, и скажем, средний ровный тон (33) в женском исполнении по частоте основного тона может оказаться высоким ровным (55) для мужчин. Не получится ли, что носители данного языка, говоря как будто об одном и том же, не смогут договориться, т.к. значения произнесенных слогов окажутся разными? Возможно, ситуация могла бы разрешиться, если бы данные регистровые тоны оценивались не в физической шкале герц, а в музыкальной шкале. Например, для мужского голоса средний тон (33) – нота «до» малой октавы, средне-высокий – нота «ми» этой же октавы, а высокий – нота «ля»; для женского – те же ноты, но уже в первой или второй октаве, будут идентифицироваться так же как мужские. Для этого надо, чтобы женские и мужские голоса различались на октаву, а все носители языка обладали абсолютным слухом.

Кроме того, если лексическое значение слога определяется его тональной характеристикой (регистром или контуром), то последняя должна быть неизменной независимо от того, как реализуется слог: с

---

<sup>3</sup> Самарина и др. Языки гэлао: Материалы к сопоставительному словарю кадайских языков / Под общей ред. И.В. Самариной.- М.: Academia, 2011.- 944 с.

данным лексическим значением - изолированно или в составе синтагмы. Но как раз это правило может не выполняться.

«... В некоторых случаях тональный контур одного и того же односложного слова при изолированном произнесении и при произнесении в составе синтагмы <...> может быть разным» [Самарина и др. 2011: 102]. Например, «при изолированном произнесении тональный контур нисходящий и изменяется от 53 до 43 St (*полутон*ов – *AB*), при произнесении в составе синтагмы нисходящий контур меняется на восходящий и реализуется в интервале 53-56 St» [Самарина и др. 2011: 102].

Столь резкое изменение тонального контура прежде всего заставляет предположить, что в составе упомянутой синтагмы слог (слово) имеет совсем не то значение, которое было реализовано информантом при изолированном произнесении.

Если лексико-грамматический анализ синтагмы всё же позволяет считать значение слога таким же как в изолированном произнесении, следует оценить, что собой представляет мелодический контур самой синтагмы в данном языке (говоре) и как он изменяет тональный контур слога. Возможно также, что «деформация» тонального контура слога является результатом влияния тональных характеристик слитно произнесенных соседних слогов. Подобная ситуация хорошо иллюстрируется приведенным ниже рисунком<sup>4</sup>.

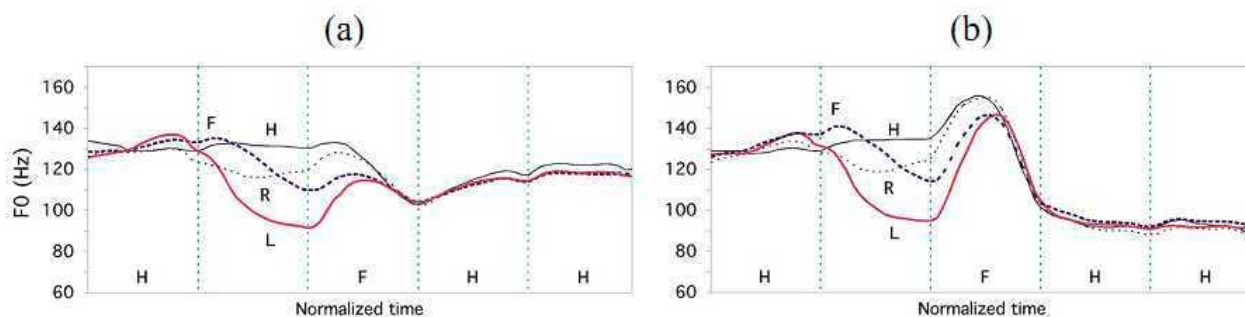


Figure 4: Mandarin tone F following four different tones. (a): no narrow focus in the sentence; (b): focus on the F-carrying syllable. Each curve is an average of 20 tokens produced by four male speakers (five repetitions per speaker). (Data from Xu 1999)

Если тональные характеристики слога действительно могут изменяться под влиянием тональных характеристик целостной синтагмы или соседних слогов, то следует допустить, что в условиях естественного речевого общения лексическое значение конкретного слога, при условии существования нескольких задаваемых тоном значений, определяется

<sup>4</sup> Xu Yi. Understanding Tone from the Perspective of Production and Perception // Language and linguistics.- 2004.- Vol.5, N 4.-P. 757-797.

Например, во вьетнамском языке постулируется наличие шести тонов, каждый из которых будучи реализован на одном и том же сегментном составе приводит к разному значению.

1. level (*không dấu*) a ă â e ê i o ô ơ u ư y ma [ mā ] = ghost
2. high rising (*dấu sắc*) á ắ ấ ế í ó ố ứ ứ ý má [ má ] = cheek
3. low/falling (*dấu huyền*) à ằ ầ ề ì ò ồ ừ ừ ỳ mà [ mà ] = but
4. dipping-rising (*dấu hỏi*) ả ẳ ẩ ể ỉ ỏ ố ử ử ỷ mà [ mǎ ] = tomb
5. high rising glottalized (*dấu ngã*) ã ắ ắ ể ể ỉ ỏ ố ử ử ỷ mã [ mã' ] = horse
6. low glottalized (*dấu nặng*) a ă â e ê i o ô ơ u ư y ma [ mà' ] = rice seedling

Если же фонетист-исследователь просит назвать тот или иной предмет (скажем «лошадь» или «щека»), носитель вьетнамского языка постарается использовать имеющиеся у него тональные средства, чтобы произнести именно это слово (именно с таким значением), но в контексте фразы эти тональные изменения могут оказаться совершенно несущественными.

И вот пример этих тональных изменений: шесть тонов вьетнамского языка в изолированно произнесенных слогах. Особого внимания заслуживает длительность этих тонов - пятьсот миллисекунд. Ни в каком речевом сигнале гласных длительностью пятьсот миллисекунд практически не бывает, это в лучшем случае длительность слога, а иногда целого слова, целой словоформы.

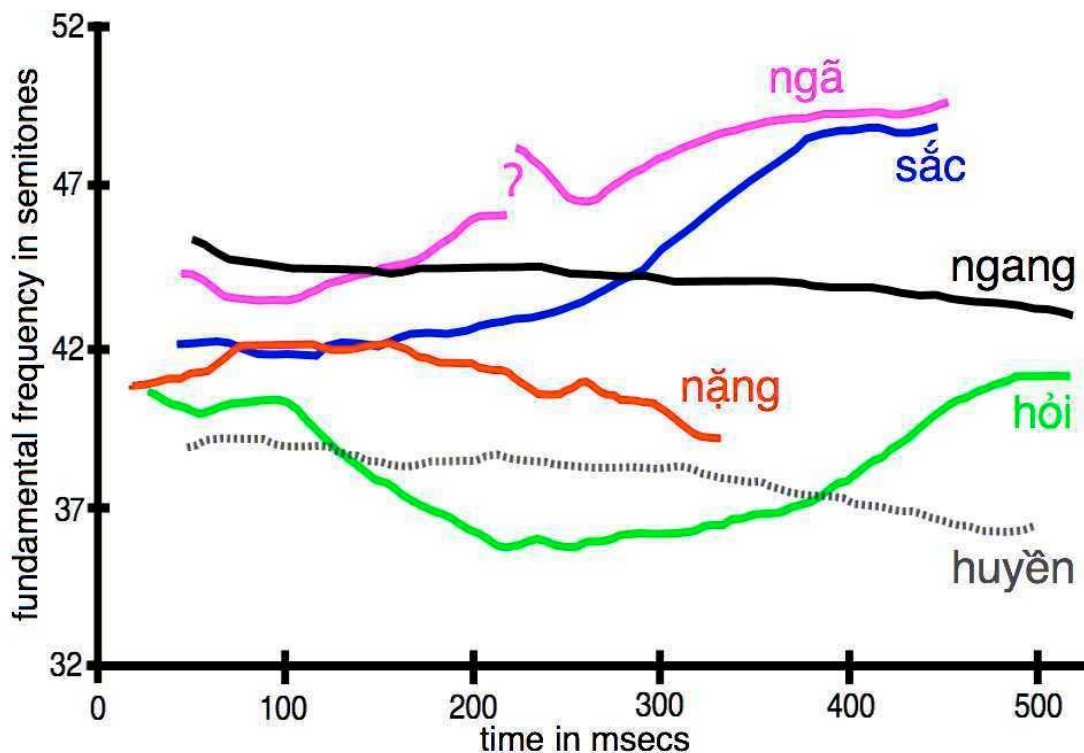


Рис. 21.

В исследованиях возможностей системы речеобразования породить быстрые мелодические изменения обнаружилась определённая закономерность между величиной изменения частоты основного тона и необходимым для этого временем: “the relations of minimum time of pitch rises and falls as functions of pitch change size are represented by the linear equations in (3) and (4), respectively,

$$(3) \quad t = 89.6 + 8.7 \, d$$

$$(4) \quad t = 100.4 + 5.8 \, d$$

where  $t$  is the amount of time (ms) it takes to complete the pitch shift, and  $d$  is the size of pitch shift in semitone” [Xu 2004: 760].

Соответственно, при типичной длительности гласного в спонтанной речи, редко превышающей 100 мс, реализовать контуры с приведенными выше изменениями частоты основного тона практически невозможно. Ниже приведены примеры мелодического оформления фразы из прочитанного носителем языка текста на вьетнамском языке. Результаты инструментального анализа свидетельствуют, что предписываемые правилами тональные изменения могут оставаться не реализованными в естественной речи в силу краткости соответствующих сегментов (гласных).

## Các bên tham gia hợp đồng hợp

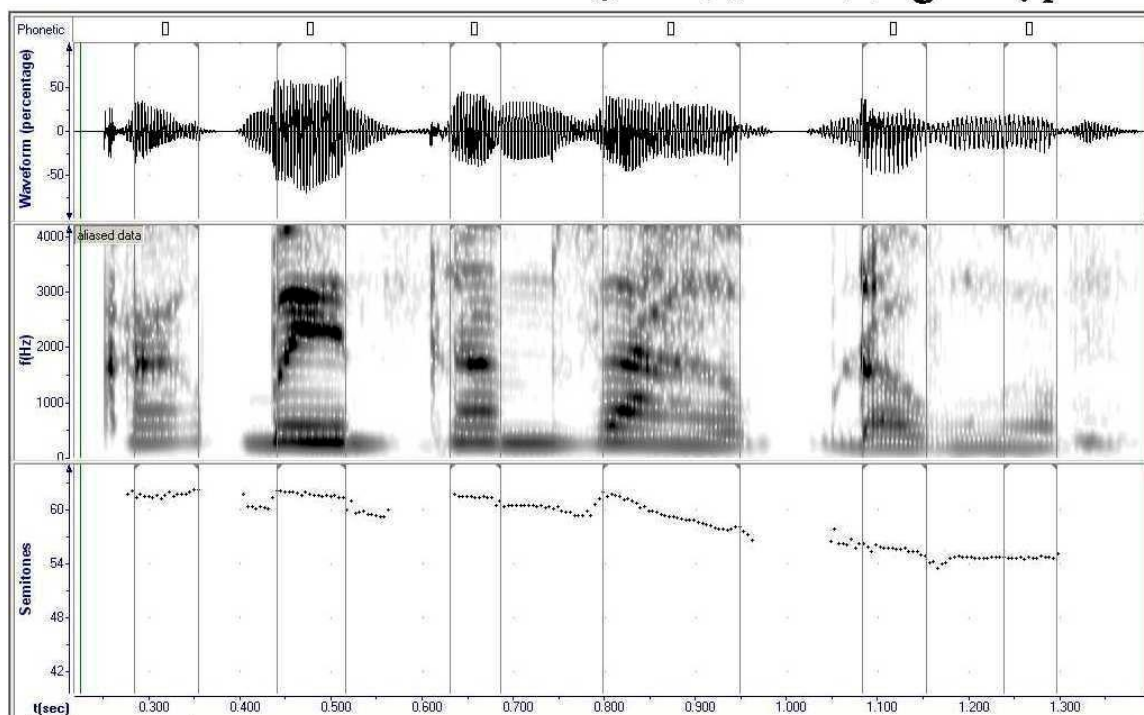


Рис. 22.

Các – по правилам вьетнамского языка: высокий повышающийся тон начинается немного выше обычного уровня речи, затем примерно на середине слога мелодия резко уходит вверх; он обозначается на письме акутом: *má, bá*. В реальности соответствующий гласный имеет длительность около 70 мс при высоком ровном основном тоне.

Получается, что в реальном (особенно в спонтанном) речевом сигнале искать тоны, характеризующие тональный язык, весьма проблематично. Конечно, когда фонетист исследует звуковой строй этого языка и просит носителя произносить заданные слоги, носитель изолированно произносит слоги или гласные необходимой длительности и со всеми нужными тональными изменениями, но когда этот же носитель будет произносить целостную фразу, то подобных тональных изменений в его речевом сигнале может не оказаться. При этом его речь остаётся понятной слушателям. Отсюда возникает вопрос: насколько необходимы «обязательные» тональные изменения?