РАЗЛОЖЕНИЕ ЗВУКОВОГО СИГНАЛА НА СОНЕЛИ

Д.т.н, профессор В.Р. Женило (Академия управления МВД РФ), М.В. Женило, к.ф.-м.н. Д.Н. Калюжный

Большинство современные исследования речевых сигналов строятся на основе их покадрового разложения на отдельные гармонические компоненты с помощью преобразования Фурье. Причем, ширина и частота следования во времени кадров спектрального описания речевых сигналов выбирается постоянной, но разной в зависимости от того, следы чего требуется выделить особенно четко – следы артикуляции или голоса.

При этом, стремясь получить как можно более детальную картину изменения частот гармонических составляющих, на которые раскладывается речевой сигнал с помощью преобразования Фурье, временное окно спектрального оценивания берется предельно малым по длине. Однако из-за проявления принципа неопределенности Гейзенберга, действующего в частотно-временном описании сигнала, это пожелание выполнимо далеко не всегда.

В конце прошлого века сформировалось целое научное направление выбора типа временного окна, которое используется для нарезки речевого сигнала на отдельные временные кадры. В каждом кадре в отдельности оценивается спектр гармонических составляющих, на которые можно разложить сигнал в этом кадре. Существует порядка сотни разных типов временных окон, которые получали имена авторов, впервые их предложивших. Среди таких имен наиболее ярким заслуженно является имя Хемминга, который не просто предложил тип временного окна, но и теоретически рассчитал, какими должны быть его параметры, чтобы неизбежно вносимые артефакты спектрального оценивания были минимальны.

В работе [1] эмпирически было показано, что из всех самых распространенных временных окон спектрального оценивания лучшим по критерию разрешающей способности по частоте и по времени является окно Гаусса. И примерно в то же самое время этот факт был теоретически доказан [2] на всем множестве допустимых временных окон спектрального оценивания.

Опираясь на это замечательное свойство временного окна Гаусса, была разработана технология [3] разложения речевых и иных звуковых сигналов на микрогармонические элементы следующего типа:

$$g(t) = \frac{A}{\sqrt{2\pi}\cdot\sigma_t} e^{-\frac{t^2}{2\sigma_t^2}} \cos(\omega t + \varphi), \qquad (1)$$

где A – амплитуда микрогармоники;

- $2\sigma_t$ - физическая длительность микрогармоники;
- ω - частота микрогармоники;
- φ - фаза микрогармоники;
- t – время.

Эта технология была реализована в виде компьютерной программы "Мастерская сигналов"[1], с помощью которой были решены некоторые научные[2] и практические задачи[3].

Фактически в этой системе весь исследуемый звуковой сигнал разлагается на отдельные микрогармонические элементы, сумма которых (по всем кадрам) позволяет синтезировать новый сигнал, весьма мало отличающийся от исходного звукового сигнала. И если исходный сигнал изначально состоит только лишь из множества амплитудно и частотно-модулированных гармоник, не приближающихся друг к другу менее чем на ширину разрешающей способности по частоте спектрального оценивания, то синтезированный сигнал оказывается почти равным исходному сигналу.

Однако, не смотря на такие ценные качества микрогармонического (1) описания звукового сигнала, в нем, все-таки, есть негативное свойство, которое вкратце можно сформулировать так – длины микрогармоник всех частот, на которые раскладывается исследуемый сигнал, равны между собой. А это означает, что микрогармоника высокой частоты может содержать в себе на два порядка больше периодов колебаний, чем микрогармоника низких частот.

При анализе обертонов голоса это свойство не является недостатком, а, наоборот, - необходимым требованием. Однако в некоторых исследованиях иногда требуется, чтобы все микрогармоники имели равное чис-

---

[1] Эта программа доступна на сайте www.zhenilo.narod.ru для использования и свободного распространения.

[2] Впервые была произведена оценка относительных мощностей периодической и апериодической составляющих вокализованных участков речевого сигнала [4]. Кроме того, аналитичность окна Гаусса позволила, использовать всю информацию о комплексном спектре, а не только о его амплитудной части. Так по одному лишь мгновенному комплексному спектру, оценивается степень амплитудной и частотной модуляции отдельных гармонических компонент, присутствующих в текущем кадре исследуемого звукового сигнала.

[3] С помощи программы "Мастерская сигналов" возможно выделение отдельных гармонических составляющих сигнала, например, обертонов голоса, скрипа двери, мощнейших фоновых помех и т.п. Выделенные следы звука можно прослушивать по отдельности, усиливать, ослаблять или, вообще, удалять. Это нашло применение в учебном процессе, когда учащиеся осваивают законы передачи информации с помощью звука.

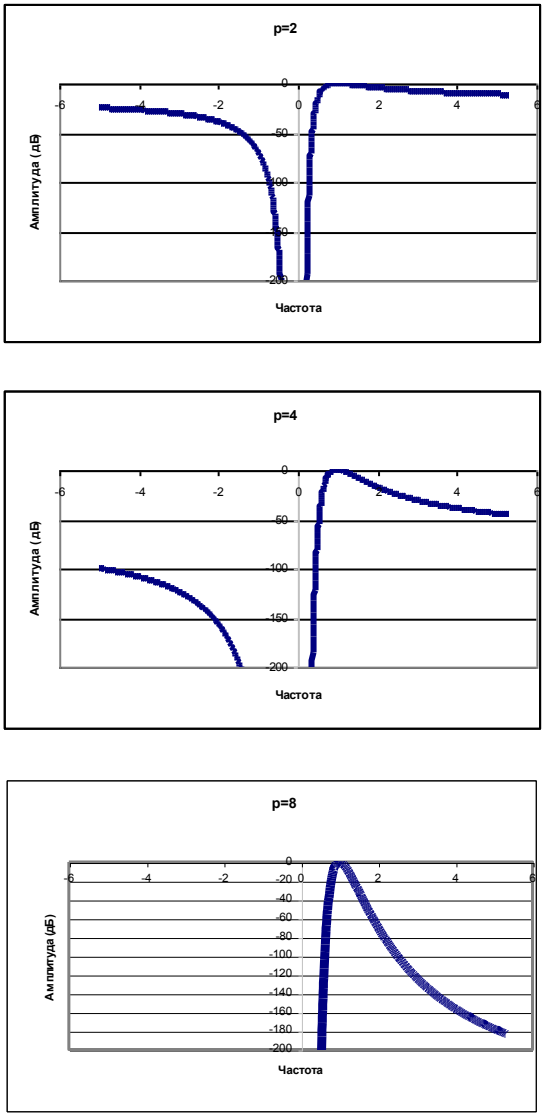

Рис.1. Амплитудные спектры преобразования Фурье-Гаусса (2) одиночной гармоники для разных значений параметра *p*.

ло периодов колебаний. В этом случае временное окно спектрального оценивания (1) должно иметь не одну и ту же ширину, равную $2\sigma_t$, для всех частот, а – разную.

Обычный динамический комплексный спектр $S(t,\omega)$ сигнала $s(t)$, лежащий в основе традиционной сонограммы, имеет вид:

$$S(t,\omega)=\int_{-\infty}^{\infty} s(t+\tau)\frac{1}{\sqrt{2\pi}\cdot\sigma_t}e^{-\frac{\tau^2}{2\sigma_t^2}}e^{-i\omega\tau}d\tau \quad (2)$$

.

По сути своей – это обычная свертка сигнала с микрогармоникой (1) единичной амплитуды.

Предлагаемое ниже преобразование Фурье-Гаусса является частным случаем обобщенного вейвлет преобразования. Смысл выделения такой частности из общего будет пояснен ниже.

Оставляя количество периодов, попадаемых в одно временное окно, в качестве пока что неопределенного параметра $p=\sigma_t\omega$, разложение (2) на микрогармоники переменной длины (но одинакового числа их периодов колебания) можно представить в следующем виде:

$$S(t,\omega)=\frac{\omega}{\sqrt{2\pi}\,p}\int_{-\infty}^{+\infty} s(t+\tau)\cdot e^{-\frac{\omega^2\tau^2}{2p^2}-i\omega\tau}d\tau. \quad (3)$$

Именно это преобразование в дальнейшем мы будем называть преобразованием Фурье-Гаусса. Некоторые свойства этого преобразования были описаны ранее [5]. Здесь же приведем лишь традиционное описание свойства этого преобразования на примере амплитудного спектра одиночной гармоники некоторой частоты F=1. Для этого на рис. 1 показаны амплитудные спектры, полученные с помощью преобразования Фурье-Гаусса, при значениях параметра *p* = 2, 4, 8.

Из этих рисунков видно, что при значении параметра *p* более, чем 8, при использовании преобразования Фурье-Гаусса вообще не возникает никаких проблем, связанных с появлением, так называемых, отрицательных частот, которые обычно имеют место при простом преобразовании Фурье. И не возникают артефакты в виде локальных максимумов, указывающих на следы ложных гармоник на уровне –200 дБ, каковые имеются у большинства традиционных взвешивающих временных окон спектрального оценивания.

Другой особенностью преобразования Фурье-Гаусса является то, что разрешающая способность спектрального представления (3) ухудшается по частоте по мере перехода от низких - к высоким частотам. Но из (3) видно, что при этом на высоких частотах повышается (в полном соответствии с принципом неопределенности Гейзенберга) разрешающая способность по времени.

Это хорошо известный факт в теории вейвлет преобразований. Но поскольку вейвлет преобразования оперируют с функциями более широкого класса со специально оговоренными свойствами [2], то мы ограничимся одним, но оптимальным типом временного окна – окном Гаусса. Обоснованием этого шага являются следующие соображения.

Весь спектральный анализ звуковых сигналов строится на базе разложения исследуемого сигнала на гармоники[4] с разными частотами. Делается это потому, что в природе звука есть одно свойство, благодаря которому удается передавать информацию, несмотря на то, что сам звуковой сигнал весьма значительно меняет свою форму при распространении его сквозь разные передающие звенья (среды). А именно: в природе существует лишь один тип простейшего сигнала, который никогда не меняет свой структуры (формы) при прохождении через любые линейные системы передачи звука. Этим сигналом является гармоника с некоторой частотой,

[4] А не на иные базовые функции, например, - степенные функции в преобразовании Тейлора.

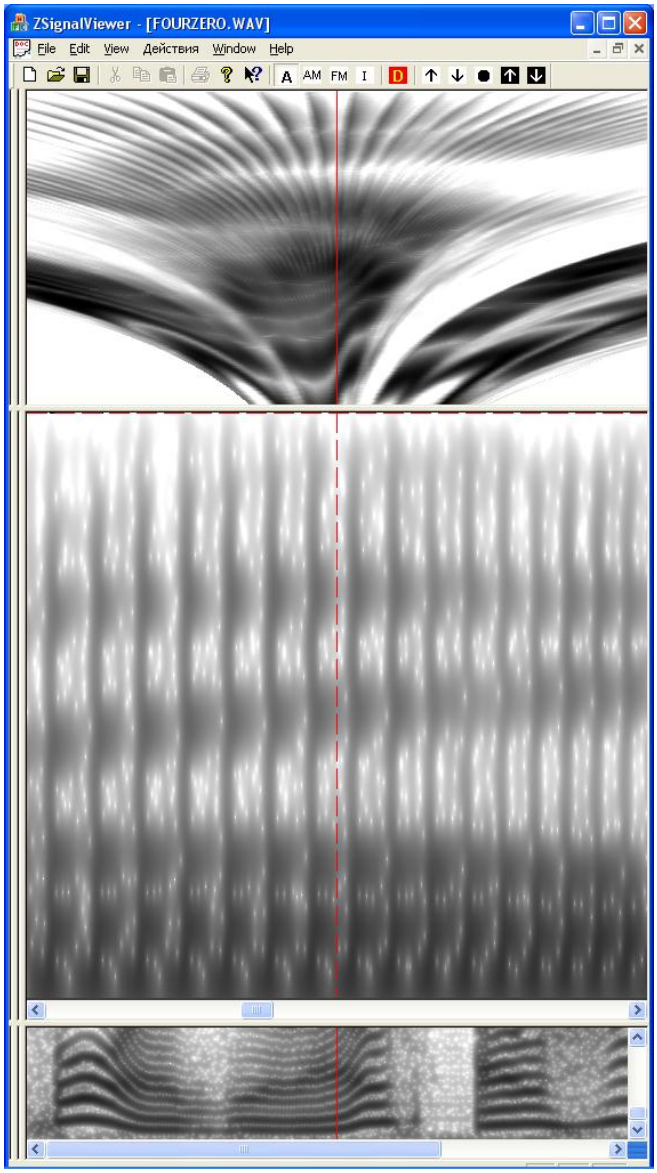

Рис.2. Три формы частотно-временного амплитудного описания речевого сигнала (пояснения в тексте).

амплитудой и фазой. И у этого сигнала есть один параметр, инвариантный к любым трансформациям, которым он подвергается, при прохождении через любые линейные системы передачи звука. Этот параметр – частота гармоники.

Именно благодаря этому свойству удается передавать информацию с помощью звука. И в этой связи, двухсотлетней давности гениальная находка Жана Батиста Жозефа Фурье просто бесценна для понимания и описания передачи информации с помощью звуковых сигналов.

Чтобы не путать микрогармоники (1) с базовыми функциями, используемыми в преобразовании Фурье-Гаусса (3), последние предлагается назвать **сонелями**. Этот термин, возможно, поможет более четко понимать и формулировать свойство тех неделимых (атомарных) информационных объектов, на которые раскладывается любой звуковой сигнал. И особо подчеркнем - эти объекты (сонели) обладают тем же важным свойством, которым обладают все гармонические звуковые сигнала – частота их колебания не изменяется при прохождении через любые линейные передающие системы[5].

Введение нового термина – сонель резонно еще и потому, что в цифровом представлении растровых графических образов признан такой атомарный информационный объект как пиксель, условно обозначающий точечный элемент картинки. Аналогией ему в звуке может служить сонель, определяемый с помощью (3). Но не следует забывать, что сонель – это частный случай более общего понятия – вейвлет.

На рис. 2 приведены три формы частотно-временного амплитудного описания речевого сигнала с помощью сонелей и обычных микрогармоник:

- с помощью сонелей, рассчитанных по формуле (2) со значением параметра $p = 8$ (верхняя часть рисунка);
- с помощью микрогармоник относительно малой длительности, рассчитанных по формуле (1) (средняя часть рисунка – широкополосная соно-

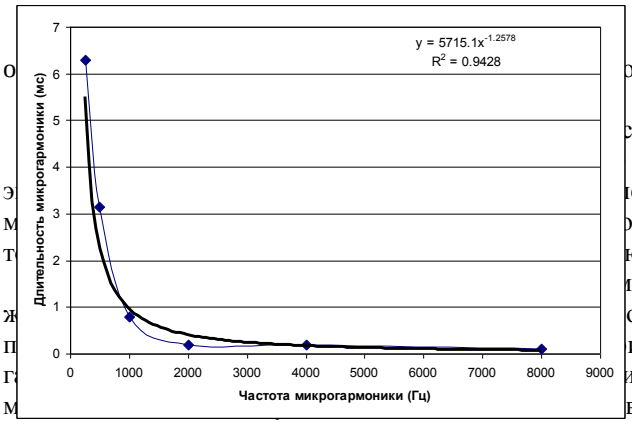

Рис. 3. Минимальная длительность микрогармоники, необходимая для точного определения на слух ее частоты.

о ой длительности, рассчитанных по формуле (1) (нижняя

сит от вопросов исследования звукового сигнала.

качеству нашему слуховому восприятию, был поставлен э ледующим образом. Какой минимальной длительности T_0 м оторой (частоту F_0) можно оценивать на слух достаточно т ющем.

икрогармоник разной частоты и длительности, и для ка- ж онику с такой же частотой. Надо отметить, что такое ис- п имента. В основном из-за того, что долго слушать чистые г их частоты под высоту звучания предлагаемых микрогар- м вания у некоторых испытуемых, настойчиво доводивших дело до конца, к сожалению, начинала болеть голова.

[5] Вообще-то, это фундаментальное свойство следовало бы во всех учебниках по акустике или передачи сигналов явно называть законом природы – законом сохранения частоты гармонического колебания, как это принято в физике, когда говорят о подобных законах сохранения материи, энергии, импульса и момента импульса и т.п.

Лучший результат из всех испытуемых показал профессиональный реставратор фонограмм, привыкший на слух находить и инструментально удалять всякие помехи, в частности, тонального типа. Один из результатов эксперимента с его участием показан на рис. 3.

В результате обработки всех полученных результатов экспериментов, выяснилось следующее. Точность определения высоты звучания микрогармоники зависит как от самой ее частоты, так и от длительности звучания микрогармоники. Общий вывод таков – чем выше частота микрогармоники, тем короче может быть ее длительность для того, чтобы с одинаковой относительной погрешностью для всех случаев установить на слух высоту ее звучания. И эта зависимость достаточно хорошо описывается следующим образом: $F_0 \cdot T_0 \approx 6$, где F_0 – это частота микрогармоники в герцах, а T_0 – длительность ее звучания в секундах.

Полученный результат подтверждает обоснованность использования преобразования Фурье-Гаусса для визуализации следов звука, однако ставит новый вопрос – как слуховой системе человека удается получить такой высокий показатель точности оценки высоты звучания сонелей, в то время как инструментальное представление с аналогичными значениями параметра *p* оказываются существенно менее выразительными, чем ожидалось.

Литература

1. *Женило В.Р. Компьютерная фоноскопия. М.: Академия МВД России, 1995. 208 с.*
2. *К. Чуи. Введение в вейвлеты: Пер. с англ. – М.: Мир, 2001. – 412 с.*
3. *Женило В.Р. Фазово-амплитудный спектральный фильм. Сборник трудов XI сессии Российского акустического общества (19-23 ноября 2001 г.). Том 3, "Акустика речи. Медицинская и биологическая акустика", М.: РАО, 2001, сс. 30-34.*
4. *Женило В.Р. Исследование соотношения периодической и апериодической составляющей речевого сигнала. Сборник трудов XI сессии Российского акустического общества (19-23 ноября 2001 г.). Том 3, "Акустика речи. Медицинская и биологическая акустика", М.: РАО, 2001, сс. 34-38.*
5. *Женило В.Р. Информация, звук и преобразование Фурье-Гаусса. Сб. трудов XV Международной научной конференции "Информатизация и информационная безопасность правоохранительных органов", М.: Академия управления МВД России, 2006, сс. 332-338.*
6. *Женило В.Р., Женило М.В., Калюжный Д.Н. О точности слухового восприятия гаусс-гармоники. Сб. трудов XIV Международной научной конференции "Информатизация и информационная безопасность правоохранительных органов", М.: Академия управления МВД России, 2005, сс. 354-356.*