

ИНФОРМАЦИЯ И ЗВУК

Д.т.н., профессор В.Р. Женило (Московский государственный лингвистический университет)

В вышедшей в свет в этом году монографии [Потапова Р.К. 2006] во введении перечисляются актуальные вопросы развития современной прикладной фонетики. Среди них вопросы развития и формирования новых технологий исследования речи, в частности, для решения задач диагностики и идентификации личности по речи имеют большое научное и прикладное значение. То, что эти вопросы чрезвычайно сложны, говорит уже тот факт, что поиск их решения ведется всем ученым миром уже более полувека. Здесь невольно вспоминается одна замечательная реплика Г.Фанта, который в ответ на замечание – компьютеры развиваются быстрыми темпами, когда же будут созданы автоматические системы распознавания речи – ответил, что для этого потребуются не компьютеры пятого поколения, а пятое поколение ученых, занимающихся решением этой проблемы.

В нашей стране первые поколения таких ученых формировались, в частности, в ходе регулярных встреч на школах-семинарах, получивших название «Автоматическое распознавание слуховых образов» (знаменитые АРСО). Это была на самом деле прекрасная школа отечественных речевиков. И, несмотря на то, что в научных дискуссиях принимали участие специалисты разных областей знаний, говорящих на разных научных языках, шло постоянное проникновение достижений из одной научной области в другую, что в итоге привело к большим успехам в теории и практике исследования речевых сигналов. Позже эти достижения стали активно использоваться на практике. Правоохранительные органы стали создавать свои курсы подготовки экспертов-фоноскопистов, поскольку таких широкопрофильных специалистов, в просторечии называемых речевиками, не готовило ни одно учебное заведение страны. Надо признать, что и до сих пор официально не существует специальности, занимающейся вопросами диагностики и идентификации личности по речевому сигналу. Из-за этого до сих пор экспертиза по идентификации личности по речи производится не одним экспертом речевиком, а, как правило, двумя экспертами – лингвистом и акустиком. Исторически сложилось именно так, хотя, непонятно почему именно ими, а не фониатрами, логопедами, психологами, психиатрами, невропатологами или специалистами с хорошим фонетическим, музыкальным, эмоциональным или индивидуальным слухом. Но, так или иначе, уже давно в этой пограничной научной области происходит взаимное проникновение научных понятий и выработка новых общих, отвечающих традициям и требованиям каждой научной области, которые представляют не только гуманитарии – лингвисты, психологи, социологи и философы, но и естественники – физики, акустики, электронщики и инженеры.

Уже несколько лет в Московском государственном лингвистическом университете на кафедре прикладной и экспериментальной лингвистики под руководством Родмонги Кондратьевны Потаповой проводятся курсы подготовки экспертов-фоноскопистов для правоохранительных органов. Тематика курса такова: «Основы прикладного речеведения применительно к лингвокриминалистике». На этих курсах, да и не только на них, но и на занятиях по информатике со студентами этого же университета рассматриваются сложные теоретические вопросы, взгляды на которые в разных научных областях несколько различаются. Среди них, пожалуй, самым спорным основным и сложным является вопрос о том, что такое информация и как ее извлекать из исследуемого речевого сигнала.

Очень часто на занятиях или на экзаменах можно наблюдать такую картину. Вне зависимости от типа высшего базового образования (гуманитарного или естественного) большинство слушателей указанного курса, студентов или аспирантов на вопрос – что такое информация, зачастую дают примерно одинаковые ответы, в которых сами же начинают сомневаться и улавливать алогизмы после некоторых уточняющих вопросов экзаменатора. Например, на вопрос – что такое информация и в чем она измеряется, могут ответить:

- в буквах, в страницах, в печатных листах, если это письменный текст;
- в секундах или часах, если это устная речь;
- в байтах или в битах, если сообщение передается в виде файла;
- и даже в рублях или в других видах валюты, если это «очень ценная информация».

Почему такое бывает понять не сложно, если учесть два решающих обстоятельства. Во-первых, как это ни странно, но найти общий и приемлемый для всех учебник по теории информации практически невозможно. Есть хорошие единичные учебники с таким названием, но они, как правило, посвящены проблемам естественных наук или проблемам передачи информации с помощью разного рода сигналов. Использовать такие учебники в гуманитарных учебных заведениях никто не собирается, даже, несмотря на то, что в некоторых дисциплинах, например, в курсах криминалистики понятие информации и ее измерение является ключевым.

Вторая причина лежит вообще вне сферы науки или учебного процесса, и корни ее уходят, по-видимому, в сферу, близкую к, так называемым, «средствам массовой информации». Именно они каждый день и каждый час прививают всем нам неправильное понимание смысла – что такое информация. В выпусках новостей, передаваемых по телевидению или радио можно услышать фразы: «Передаем информацию» или «Послушайте информацию». Если на других радиостанциях могут сказать: «По поступившим сообщениям» или «По имеющимся сведениям или данным», что достаточно корректно,

поскольку их истинность далеко не всегда оказывается известной в момент их изложения, в силу высокой оперативности новостных технологий, то другие, предпочитают обороты: «Информация о заложенной бомбе не подтвердилась», «За достоверную информацию гарантируется вознаграждение», «С места происшествия поступает противоречивая информация» и т.п. Даже в московском метро несколько лет назад можно было услышать фразу (из «средств местного массового оповещения пассажиров»): «Уважаемые москвичи и гости столицы, послушайте, пожалуйста, информацию о правилах поведения в метрополитене». Но, справедливости ради, надо отметить, что в таком виде это сообщение просуществовало не очень долго. В конце концов, его изменили, выбросив слово «информацию». Но на смену этой ошибки пришла новая. Сейчас очень часто в метро можно услышать предупреждение о том, чтобы пассажиры не доверяли самовольно расклеиваемым объявлениям, поскольку «... в них может содержаться заведомо ложная информация».

Невдумчивое или поверхностное отношение к важнейшему ключевому понятию современной науки (да и жизни вообще) - информации, глубоко въелось в наше обыденное сознание. Глубину этого явления можно оценить еще и по тому, что мы настолько привыкли к некоторым словосочетаниям, подобным СМИ, что они давно уже встречаются не только в быту, но и в научных публикациях. Например, можно встретить словосочетание - «обработка информации», а не обработка сообщений, данных или записей. А в специальностях ВАК находим - «защиту информации», вместо защиты сообщений, носителей, каналов, данных и т.п.

В разных словарях, справочниках и учебниках можно найти не менее десятка принципиально разных определений понятия «информация». Приведем лишь пять из них, которые наиболее характерно выделяются среди других, и, по сути, впитывают в себя смысл остальных.

Начнем с классического «Словаря русского языка» Сергея Ивановича Ожегова. Как и все словари подобного рода, в нем практически для каждого русского слова дается его толкование или определение. И естественно, в таком словаре мы очень быстро находим замкнутую цепочку, по которой можно кружить бесконечно долго, поскольку в таких словарях нет первичных понятий, не определяемых ни через какие иные простейшие понятия. Поэтому, начав с «информации», мы можем перейти к «сведениям». От «сведений» - к «знаниям». И потом от «знаний» снова вернуться к «сведениям».

Это типичная ситуация «зацикливания» для подобного рода справочников, в которых отсутствует ряд изначальных (первичных) понятий, которые не определяются ни через какие иные. Для живого русского языка – это нормальное явление. Но в науке, требующей строгости и, в первую очередь, однозначности, так не принято. Хотя, такое иногда встречается. Забегая вперед, сразу отметим, что, не смотря на то, что основные элементы теории информации были заложены Клодом Элвудом Шенноном в 40-е годы 20-го века, до сих пор дать четкое определение тому, что такое «информация» никто не может. И это нормальное явление в науке. Аналогичные явления происходили ранее с основными понятиями во многих научных областях.

Упомянув Шеннона, сразу же приведем его определение понятия «информации», которое считается лучшим в кругу ученых, занимающихся естественными науками: «Информация – это то, что снижает неопределенность».

В словаре по информатике, составленном Дмитрием Александровичем Поспеловым [Информатика 1994 : 41] находим такое определение: «Информация – это содержание сообщения, сигнала, памяти, а также сведения, содержащиеся в сообщении, сигнале или памяти». Вторая часть этого определения повторяет соответствующий раздел словаря С.И.Ожегова. А первая часть, говорящая о том, что «информация – это содержание сообщения ...», это весьма существенная добавка для цифровых информационных технологий, поскольку, например, если в файле содержатся некоторые сведения, то понять их невозможно, не зная, что обозначают байты этого файла. Поэтому, чтобы содержимое файла правильно интерпретировалось, сейчас к имени файла обязательно добавляется его тип.

Приведем одно из лучших философских определений понятия «информация», данное Дж. Баттесоном. Это определение, также как и приведенное выше определение, данное Шенноном, можно найти в интернете в Web-Encyclopedia. «Информация – это то, что изменяет нас». Это определение поначалу кажется правильным. Но в нем, как и в большинстве философских трактовок фундаментальных понятий, скрываются принципиальные противоречия. Например, так же как в одну и ту же философскую реку нельзя войти дважды, так оказывается, что и письмо информации не содержит до тех пор, пока оно не вскрыто и не прочитано нами. А раз оно информации не содержит, то и вскрывать его нечего.

Из всего множества определений понятия информации особое внимание привлекает к себе такое, в котором авторы не мудрствуют лукаво, а честно признаются в том, что они не могут четко сказать - что такое информация. В одном словаре авторы написали так: «Информация (от лат. informatio – разъяснение, изложение) – основное понятие кибернетики». И точка.

Поначалу может показаться странным, что такое определение дает самая щепетильная в вопросах точности своих утверждений научная отрасль. Ведь это определение приведено в математическом энциклопедическом словаре [МЭС 1998 : 245]. Но если призадуматься, то ничего странного в этом нет. И это очень хорошо объясняется далее в той же рубрике словаря. Далее в ней идут лишь разъяснения, но не определения. Однако и в этих разъяснениях говорится о том, что современное представление понятия «информации» пока еще интуитивно, и в науке оно еще окончательно не сформировано. В этой же статье

приводятся интересные аналогии из других научных областей, где также существуют ключевые понятия, которые нельзя определить через иные – более простые понятия. Например, дать определение – что такое время, масса, энергия, импульс, площадь или объем никто не может. Но, тем не менее, человечество успешно пользуется этими понятиями и, более того, умеет измерять их значения количественно. И это притом, что в природе таких эталонов и единиц измерения как секунда, грамм, метр, квадратный метр не существует. Все эти единицы измерения придумал человек¹ для удобства хранения или передачи сведений о тех или иных объектах или явлениях, которые он использует в своей жизни.

Надо отдать должное авторам упомянутого математического энциклопедического словаря, которые не остановились только лишь на корректном и честном признании невозможности дать определение понятия «информации». Памятуя о будущем развитии науки (о молодом подрастающем поколении), авторы, все же, разъясняют современные основные тенденции в становлении понятия «информации». И в конце этого же словаря в разделе «Словарь школьной информатики» они приводят следующее определение, в значительной мере повторяющее приведенные выше определения. Вот оно:

«Информация (от лат. informatio – разъяснение, изложение) – содержание сообщения или сигнала, сведения, рассматриваемые в процессе их передачи или восприятия; одна из исходных общенаучных категорий, отражающая структуру материи и способы ее познания, несводимая к другим, более простым понятиям» [МЭС 1988 : 821].

Первая часть этого определения по сути своей совпадает с определениями С.И.Ожегова и Д.А.Поспелова. А вот вторая половина – подчеркивает важнейшую мысль о том, что понятия «информация» и «структура» очень близки друг к другу.

«Структура – [лат. structura] – взаиморасположение и связь составных частей чего-либо» [СИС 1988]. Причем, здесь под взаиморасположением не следует понимать лишь пространственное расположение, но и временное, и причинно-следственное, и любое другое функциональное или взаимное положение чего-то (как материального, так и идеального) относительно чего-то.

Объединяет эти два понятия очень многое и, в частности, такое немаловажное свойство, что и «структура» и «информация» относятся к категории идеального. Тут очень полезно привести примеры о том, что свойства того или иного вещества существенно зависят не только от того, из каких атомарных материальных объектов (кирпичиков) оно состоит, но и от того, как структурированы эти атомы-кирпичики. Поэтому одни и те же атомы углерода, структурированные по-разному, мы в целом называем графитом, углем или алмазом.

Или другой пример. Если, образно говоря, взять ящик угля, баллон с кислородом и баллон с водородом, то, по-разному структурируя атомы углерода, кислорода и водорода в молекулы, мы можем получить большинство всех известных органических веществ. И поэтому вопрос о том, что важнее – материальное (атомы) или идеальное (структура) становится просто некорректным. Поскольку не бывает структуры без материи и не бывает материи без структуры.

Затронув большой философский вопрос – а что же такое материальное или идеальное, автор не претендует на его даже частичное решение, поскольку дать определение и этим первичным понятиям тоже, видимо, нельзя. Но чтобы, хотя бы, отличить материальное от идеального, можно было бы принять следующую рабочую гипотезу: то, для чего применима обычная алгебра, придуманная человечеством, в которой $A+A=2A$, то – материально. А то, для чего применима другая алгебра – алгебра теории множеств, в которой $A+A=A$, то – идеально. Здесь невольно вспоминается высказывание Джорджа Бернарда Шоу: "Если я с кем-то поделюсь яблоком, то мне достанется половина и кому-то - половина. А если я поделюсь идеей, то и у меня будет идея, и еще у кого-то".

По этому простому косвенному признаку информация относится к категории идеального. А все сообщения, несущие в себе информацию, материальны. Поэтому два идентичных сообщения в сумме дадут два сообщения, но суммарная информация от этого не увеличится.

По этому рабочему определению понятие «структура» тоже относится к категории идеального. Это возможно благодаря тому, что в определении понятия множества лежит понятие «свойство». Теория множеств – учение об общих свойствах множеств, преимущественно бесконечных. Понятие множества, или совокупности, принадлежит к числу простейших математических понятий; оно не определяется, но может быть пояснено при помощи примеров. Так, можно говорить о множестве всех книг, составляющих данную библиотеку, множестве всех точек данной линии. Книги данной библиотеки, точки данной линии являются элементами соответствующего множества. Чтобы определить множество, достаточно указать характеристические свойства элементов, т.е. такое свойство, которым обладают все элементы этого множества и только они. Может случиться, что данным свойством не обладает вообще ни один предмет; тогда говорят, что это свойство определяет пустое множество» [МЭС 1988 : 380].

Определенное расположение чего-то относительно друг друга (т.е. структура) это и есть свойство в широком смысле. Поэтому информацию и структуру нельзя обрабатывать. А если это делать, то это будет уже не та структура и не та информация.

В развитие этой темы уместно упомянуть еще об одной алгебре – алгебре комплексных чисел. Эту третью алгебру – алгебру комплексных чисел придумали для описания реальных природных явлений. Но в

¹ Здесь уместно вспомнить о том, откуда произошли единицы измерения: секунда, дюйм, вершок, метр, грамм и др.

быту эта алгебра никому не нужна. Видимо, отсюда и все сложности ее преподавания в общеобразовательных школах. Не каждый педагог может растолковать смысл этой алгебры.

Объяснять слушателям, что такое комплексные числа, видимо, не стоит. Но попытаться пояснить, почему человечество придумало, так называемые, мнимые числа, расширяющие реальные числа до комплексных, полезно. Здесь следует признать, что название «мнимое число» - не вполне удачно. Возможно, их правильнее было бы называть «скрытыми числами». Тогда получается, что комплексное число - это реальное число, которое мы привыкли «видеть или трогать» плюс тоже реальное число, но скрытое (но никак не мнимое). С помощью комплексных чисел хорошо описываются многие природные явления, но которые наши органы чувств плохо воспринимают или не воспринимают вообще. Например, если к некоторой точке тела приложить одну силу и прибавить к ней такую же, но направленную в противоположную сторону, то получится нулевая сила. Взаимодействия этих сил описываются с помощью обычных реальных чисел и для них применима обычная алгебра (и векторная алгебра). Это мы можем увидеть или ощутить. А вот почему две встречные волны на поверхности воды (или две звуковые волны) взаимно не уничтожаются при встрече друг с другом, а проходят друг через друга, не меняя своей формы (структуры, и поэтому, может быть, информации), - это скрытые от наших глаз законы природы. Но все колебательные процессы разной природы точно описываются именно с помощью комплексных чисел. Эти числа такие же реальные как реальные описываемые процессы. Поэтому название «мнимое» число - это такой же алогизм, как название «средства массовой информации».

Ещё одним важным свойством «структуры» и «информации» является то, что они не могут быть истинными или ложными. Если хотите, то структура всегда истинна. Она ложной быть не может. Структура - вот она. Какая она есть, такая она и есть. А вот то, что кто-то может утверждать, что данная структура совпадает или нет с другой структурой - это уже совсем другой вопрос. То же самое и с «информацией». Видимо, не зря человечество для понятия, противоположному «информация», употребляет термин - «дезинформация» (которая всегда ложна). Не может быть дезинформация истинной, а информация ложной. Поэтому если кто-то говорит: «Послушайте информацию» (а не сообщение), - то на подсознательном уровне это может восприниматься так: «Послушайте правду». Ведь никто никогда не скажет: «Послушайте дезинформацию».

Приведенные соображения, конечно же, не всеми будут признаны справедливыми. И это естественно. Каждый в своей научной сфере привык видеть изложенное выше под своим углом зрения. А граней у этих понятий так много, что единое мнение ещё не скоро сформируется. Например, в широких кругах уже давно витает идея о том, что информацию надо различать по двум типам, как это, например, принято в физике: потенциальная и кинетическая. Потенциальная информация² хорошо оценивается количественно с помощью энтропии³, что позволило Клоду Элвуду Шеннону осуществить величайший прорыв в теории передачи информации. Как понимать «кинетическую» или «не потенциальную» информацию пока ещё не ясно.

Заканчивая столь длинное философское отступление от главной темы изложения, перейдем к основному вопросу - что в звуке можно назвать информацией?

Следуя изложенным выше принципам, наверное, можно было бы утверждать, что информация в звуке - это структура самой звуковой волны. И не важно как эта структура представлена, и в какой системе координат: «время», «пространство», «давление звукопередающей среды». Или в системе координат: «время», «напряжение электроакустического сигнала». Или в какой-то другой системе координат.

Но хорошо известно, что один и тот же исходный речевой сигнал, записанный с помощью нескольких микрофонов и разной электроакустической аппаратуры, с применением магнитной записи или без, в итоге дает нам осциллографические описания, не тождественные друг другу. И эти отличия могут быть настолько велики, что, глядя на осциллографические описания одного и того же речевого сигнала, полученные с помощью разной аппаратуры, начинающий исследователь может посчитать, что это разные речевые сигналы. И только лишь после их прослушивания, он понимает, что это один и тот же речевой сигнал, структура которого до неузнаваемости изменилась из-за того, что исходный речевой сигнал прошел через разные электроакустические звенья и зафиксировался на разных носителях информации.

Любому начинающему речевнику такие метаморфозы речевого сигнала могут показаться весьма странными и непонятными. Как это так - формы (структуры) осциллограмм речевых сигналов на приборе или компьютере совершенно разные, а на слух четко определяется, что это один и тот же речевой сигнал?

Разбираясь с этим загадочным явлением, любознательный речевик со временем начинает понимать главное. Оказывается, что, благодаря гениальной находке Жана Батиста Жозефа Фурье, которой уже 200 лет, любой сигнал (и речевой - в том числе) можно разложить на атомарные информационные элементы, некоторые параметры которых несколько не меняются (остаются инвариантными) при их прохождении через любые естественные системы передачи звуковых сигналов. Этими «атомарными» элементами являются гармоники, которые описываются всего лишь тремя параметрами: амплитудой, частотой и начальной фазой колебания.

² Термин потенциальности и кинетичности здесь употребляется не в качестве предложения, а в качестве иллюстрации и толкования мысли о двух возможных видах информации.

³ Понятие энтропии ввел физик Рудольф Юлиус Эмануэль Клаузиус в 1865 году.

Несмотря на конечную продолжительность во времени любого речевого сигнала, он однозначно и с абсолютной точностью представим с помощью преобразования Фурье в виде суммы гармоник разных частот, амплитуд и фаз. И эти гармоники, вообще-то, имеют бесконечную продолжительность. Последнее обстоятельство очень непривычно для начинающего речевика. Но, преодолев этот, в некотором роде, психологический барьер, он получает в свои руки незаменимый инструмент извлечения из речевого сигнала всей интересующей его информации. И это, не смотря на то, что осциллографическая форма речевого сигнала может изменяться до неузнаваемости в процессе его прохождения через разные акустические и электроакустические среды.

В чем же заключается ценность разложения любого речевого сигнала на спектр гармоник и какая связь всего этого с информацией и структурой?

К сожалению, в большинстве учебной, да и научной литературе для начинающих речевиков обычно не подчеркивается – почему (для чего) речевой сигнал раскладывается именно в спектр гармоник с помощью преобразования Фурье, а не в спектр каких-либо иных базовых функций, например, степенных, как это делается в разложении любой функции в ряд Тейлора.

Объяснить это можно следующим образом. Дело в том, что простейший звуковой сигнал, представленный в виде кривой гармонического типа, обладает одним очень ценным природным свойством, позволяющим использовать его для передачи информации. А именно, из всех трех параметров, описывающих любую гармонику, есть лишь один такой, значение которого остается инвариантным (неизменным) к любым трансформациям, претерпеваемым звуковыми сигналами, после их прохождения через любые естественные (и большинство искусственных) каналов передачи. Этим параметром является частота гармоники. Значения же остальных двух параметров могут значительно изменяться. Например, амплитуда гармоники, значение которой на слух воспринимается как громкость, может существенно уменьшиться или увеличиться. А значение начальной фазы колебания, показывающее положение локального максимума гармоники относительно некоторого момента времени, обычно называемого условным нулевым отсчетом времени, на слух нами вообще не воспринимается. Значение же параметра – частота гармоники – на слух воспринимается как высота звучания гармоники.

Этот природный закон сохранения частоты гармоники играет ключевую роль в передаче информации с помощью звуковых волн. Поэтому наша слуховая система: лучше всего ощущает малейшие изменения значения частоты гармоники (высоты звучания) и на порядок хуже ощущает изменения амплитуды гармоники (громкости звучания).

Таким образом, несмотря на то, что сама форма речевого сигнала существенно изменяется при прохождении его через разные передающие звенья, в нем, все же, можно выделить структуру некоторых элементов, которая остается инвариантной, и благодаря которой надежно передается информация.

Этой структурой является структура частот отдельных гармоник, на которые можно разложить любой звуковой сигнал в каждом отдельном кадре его покадрового описания.

Именно на этом принципе основана технология визуализация следов звука с помощью классических сонаграмм (динамических спектрограмм). Но в этой технологии есть одна сложность, заставляющая исследователей следов звука строить и анализировать не одну, а как минимум две сонаграммы. Речевики обычно называют эти два типа сонаграмм широкополосными и узкополосными. Это название закрепилось после выпуска фирмой “Kay Elemetrics Corp.” прибора визуализации следов звука - сонаграфа, в котором не было понятия «кадр звука», но было понятие «разрешающая способность по частоте» отдельных гармонических компонентов звука.

Широкополосные сонаграммы используются для лучшей визуализации следов артикуляции (резонансных частот речевого тракта), а узкополосные сонаграммы – для лучшей визуализации следов обертонов голоса. Дать четкое техническое определение того, что такое узкополосная, а что такое широкополосная сонаграммы нельзя. Это весьма размытое понятие. Так, например, одна и та же сонаграмма для низкого мужского голоса может быть широкополосной, а для высокого женского голоса – оказаться узкополосной.

В современной компьютерной (цифровой) технологии исследования речевых сигналов преобладают технологии нарезания речевых сигналов на отдельные кадры, в каждом из которых с помощью преобразования Фурье находятся те самые атомарные информационные элементы, один из параметров которых всегда остается неизменным. Этими атомарными элементами являются отдельные гармоники. А неизменными параметрами - частоты гармоник, на которые раскладывается вырезанный в кадре фрагмент речевого сигнала.

В технологии вырезания одного кадра речевого сигнала с целью определения его частотного спектра используются специальные функции, которые иногда называются весовыми, взвешивающими или оконными.

В качестве весовой функции или, иначе говоря, способа покадрового вырезания фрагмента сигнала для определения состава атомарных информационных элементов (гармоник) в нем, можно выбрать, вообще-то, любую функцию. Самые распространенные способы вырезания кадра основаны на оконных функциях: прямоугольное окно, окно Хемминга, Хеннинга, Бесселя, Гаусса и некоторые другие. Такое обилие временных окон нарезания сигнала на кадры может запутать не только начинающего, но и опытного речевика. На самом деле, какому из этих окон следует отдавать предпочтение? Обычно на этот вопрос

отвечают так – все зависит от объекта и целей анализа. И это правильно. Но объектом анализа является звуковой сигнал, в котором нас интересует его инвариантная структура взаимного расположения составляющих гармоник разных частот. А раз так, то использование для этой цели прямоугольного временного окна полностью противопоказано, поскольку его применение не только не выявит всю структуру частот гармоник, но и внесет массу артефактов (ложных следов гармоник, которых реально в анализируемом сигнале нет). То же самое касается и многих других видов временных окон.

Более того, анализируя на сонаграмме следы отдельных гармонических составляющих разных частот, на которые раскладывается исследуемый сигнал, мы фактически раскладываем этот сигнал на отдельные короткие участки гармоник, по длительности и огибающей форме совпадающих со взвешивающей функцией. Иначе говоря, каждая из этих составляющих фактически является атомарным элементом звука, который до сих пор еще не имеет своего названия.

В научной литературе чаще всего эти атомарные элементы называются вейвлетами. А по сути их можно было бы назвать, например, сонелями (от sound element), копируя название пикселя (от picture element) в цифровом описании графических образов, используемом в компьютерной технологии. Ранее автор этой статьи называл эти элементы микрограмматиками и даже гауссгармониками. И приготовил компьютерную систему программ «Мастерская сигналов»⁴, с помощью которой можно раскладывать исследуемый звуковой сигнал на элементарные микрогармоники (сонели), с которыми можно выполнять операции удаления, добавления, выделения (для прослушивания), изменения амплитуды, что составляет суть редактирования основной информационной структуры звукового сигнала.

Недостатком элементарных микрогармоник, используемых в «Мастерской сигналов» является то, что их длительность оказывается одинаковой для всех частот в текущем сеансе редактирования следов звука. С информационной точки зрения – это не совсем правильно. Действительно две микрогармоники на разных частотах оказываются неравноценными по критерию количества периодов, попадающих в них. Так в одном и том же весовом окне низкочастотная микрогармоника может содержать количество периодов в десятки, а то и в сотни раз меньше, чем высокочастотные микрогармоники.

Насколько важно это с информационной точки зрения теоретически ответить пока что сложно⁵. Но следующий психоакустический эксперимент, проведенный автором с рядом добровольцев естествоиспытателей, принципиально подтвердил важность структурного спектрального описания звука с помощью микрогармоник не фиксированной, а переменной длины. Суть этого эксперимента заключалась в следующем.

Испытуемому предлагалось прослушать ряд микрогармоник разной частоты и длительности, и для каждой из них подобрать на слух долго звучащую гармонику с такой же частотой. Частоты микрогармоник выбирались случайным образом из ряда: 125, 250, 500, 1000, 2000, 4000, 8000, 16000 Гц. А длительность микрогармоник тоже случайным образом выбиралась из следующего ряда возможных значений: 12.57, 6.28, 3.14, 1.571, 0.785, 0.393, 0.1964, 0.0981 мс.

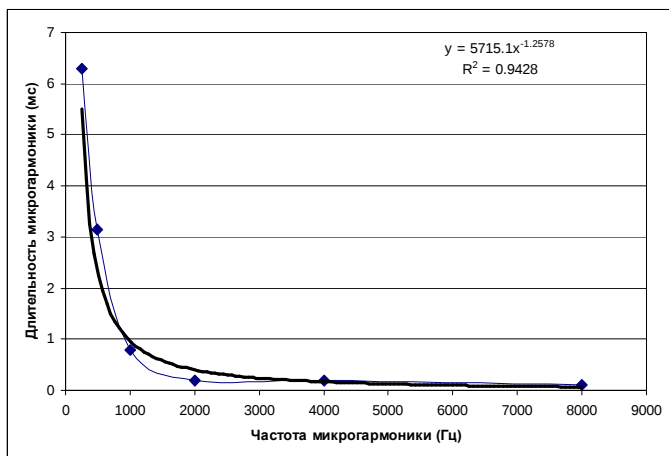


Рис. 1. Экспериментально установленная минимальная продолжительность звучания микрогармоники, необходимая для точного определения на слух ее частоты.

Всего в эксперименте с одним испытуемым использовались не более 64-х комбинаций частот и длительностей микрогармоник, в зависимости от того, услышал ли испытуемый на подготовительном этапе исследования все гармоники из ряда выше перечисленных значений⁶. Надо отметить, что такое испытание выдерживали далеко не все участники эксперимента. В основном из-за того, что долго слушать чистые гармонические сигналы в процессе подбора вручную их частоты под высоту звучания предлагаемых микрогармоник оказалось очень утомительно. В конце исследования у некоторых испытуемых, настойчиво доводивших дело до конца, к сожалению, начинала болеть голова. И как ни странно, худшие результаты⁷ показали профессиональные музыканты,

⁴ Система программ «Мастерская сигналов» открыта для широкого доступа на сайте www.zhenilo.narod.ru/sw.

⁵ Этим сейчас активно занимается теория вейвлет анализа.

⁶ Многие испытуемые на слух слышали гармонические сигналы с частотой 16000 Гц, из-за чего ряд возможных вариантов микрогармоник с разной частотой и длительностью звучания для них автоматически сокращался с 64-х до 56-ти.

⁷ Худшие по разбросу результатов отдельных испытаний.

привыкшие с высочайшей точностью определять высоту звучания натуральных музыкальных звуков, и, видимо, психологически не готовых слушать и работать с чистыми одиночными гармониками. Во всяком случае, для них этот эксперимент вызывал массу отрицательных эмоций. А лучшие показатели (и не в одной серии экспериментов) 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. 1.

После 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 (основанием этого послужили результаты аппроксимации, приведенные, например, на том же рис. 1 в правом верхнем углу): $F \cdot T \approx 6$, где F – это частота микрогармоники в герцах, а T – длительность ее звучания в секундах.

Как полученный результат можно было бы использовать в технологии отрисовки структуры атомарных компонентов звука (гармоник) с помощью компьютерной графики близкой к традиционной сонаграмме? Прежде, чем рассказать об этом, хотелось бы еще раз напомнить о проблеме выбора исследователем речи типа традиционной сонаграммы – широкополосной или узкополосной.

В идеале, для полноты изучения, исследователь следов звука должен строить несколько сонаграмм с разными разрешающими способностями по частоте. Это привело бы к тому, что традиционно плоская сонаграмма должна была бы стать объемной. Неявно это так и происходит, когда эксперт-фоноскопист при решении нескольких криминалистических задач строит отдельные слои этой объемной сонаграммы:

- нижний слой (с малым временным окном⁸) для получения широкополосных сонаграмм, хорошо отражающих следы артикуляции;
- средний слой (со средней длиной временного окна) для получения узкополосных сонаграмм, хорошо отражающих следы голоса;
- верхний слой (или даже несколько верхних слоев с большой длиной временного окна) для получения сверхузкополосных сонаграмм для обнаружения возможных следов фоновых помех, помогающих эксперту выявлять признаки монтажа фонограмм.

Эксперт вынужден это делать потому, что традиционно все современные сонаграммы строятся с постоянной длиной временного окна для всех гармонических составляющих звука – и для низкочастотных, и для высокочастотных. Однако это ограничение можно снять, используя высокие вычислительные мощности современных компьютеров. Для этого в преобразовании Фурье вместо фиксированной по ширине взвешивающей функции следует использовать окно переменной длины. Для низких частот ширина этого окна должна быть больше, а для высоких – меньше.

Для реализации этого подхода следует еще раз затронуть вопрос выбора типа временного окна.

В свое время автор эмпирически показал [Женило 1995 : 134], что из всех самых популярных в то время временных окон окно Гаусса обладает лучшей разрешающей способностью по времени и частоте. Примерно в то же время другой исследователь в своем учебном пособии по теории вейвлетов [Чуи 2001] приводит теорему, доказывающую то же самое⁹. Этот факт позволяет предложить новую форму преобразования Фурье, которое из-за замечательных свойств функции Гаусса, применяемой в этом преобразовании, следовало бы назвать *преобразованием Фурье-Гаусса*.

Мы не будем приводить здесь математическое определение преобразования Фурье-Гаусса, которое можно найти в работе [Женило 2006]. Вместо этого приведем примеры возможного разложения речевого сигнала на отдельные сонели, что позволяет, например, услышать то, что невозможно услышать с помощью других технологий обработки и анализа речевых сигналов. Для этого на рис. 2 показана узкополосная сонаграмма некоторого исследуемого речевого сигнала. А на рис. 3 показаны лишь некоторые отдельные микрогармоники (сонели), на которые раскладывается этот речевой сигнал. На рис. 4 выделены цепочки микрогармоник (сонелей), из которых состоят следы двух обертонов голоса этого же речевого сигнала. И эти следы можно раздельно прослушать. На рис. 5 показан результат сборки нового (синтезированного) речевого сигнала собранного из всех сонелей, на которые раскладывается исходный речевой сигнал, показанный на рис. 2. Звуковые файлы этих примеров и дополнительные пояснения можно найти на указанном ранее интернет сайте.

⁸ Читай – малой длительностью микрогармоник, на которые раскладывается исследуемый звуковой сигнал.

⁹ К сожалению, К. Чуи приводит эту теорему без доказательства, ссылаясь на громоздкость оного.

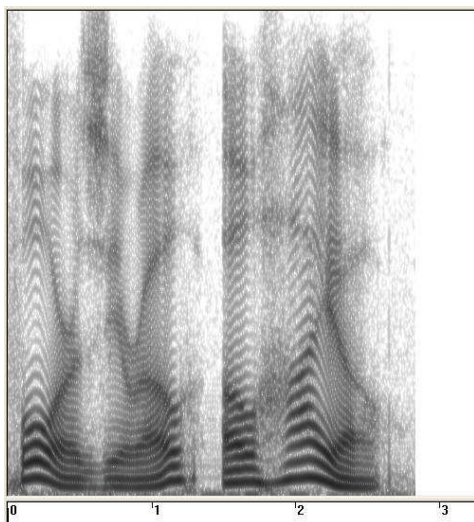


Рис. 2. Узкополосная сонаграмма речевого сигнала.

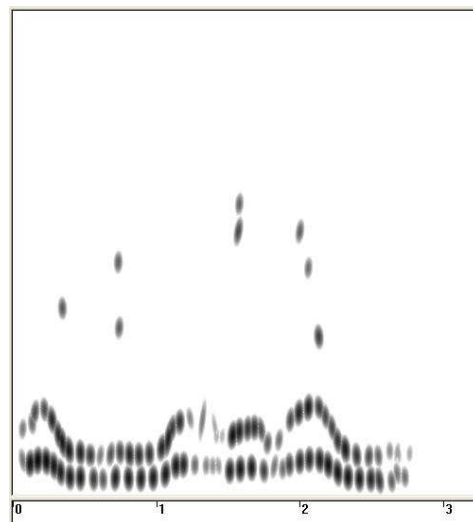


Рис. 3. Отдельные сонели, на которые раскладывается речевой сигнал с рис. 2.

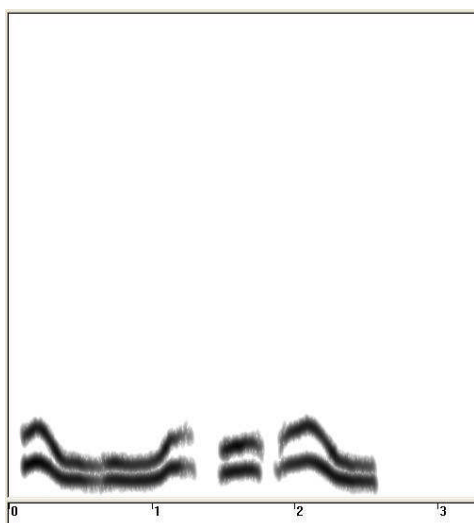


Рис. 4. Непрерывные цепочки сонелей, на которые раскладываются выделенные обертона голоса с рис. 2.

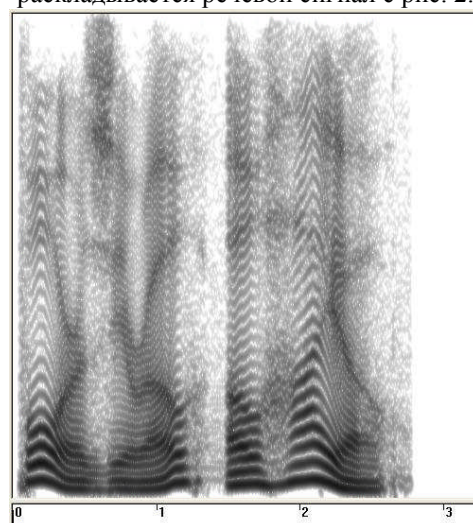


Рис. 5. Синтезированный речевой сигнал, собранный по всем сонелям, на которые раскладывается речевой сигнал с рис. 2.

В завершение повторим основную мысль всего предыдущего изложения.

Структурными атомарными инвариантными составными элементами любого звукового сигнала, в том числе и речевого, являются элементарные гармоники, на которые можно разложить любой фрагмент речевого сигнала. Структура частот этих гармоник, если она не исчезает или не разрушается полностью при прохождении речевого сигнала через каналы его передачи, является лучшим инвариантным (информационным) отражением речевого сигнала. Это неявно используется нами при слуховом восприятии и анализе звука и явно используется в инструментальных технологиях спектрального анализа речевых сигналов.

Все вышеизложенное является частью нового курса: «Основы прикладного речеведения применительно к лингвокриминалистике», - который проходят будущие специалисты по идентификации личности по речи. А подготовкой специалистов обозначенного профиля занимается кафедра прикладной и экспериментальной лингвистики Московского государственного университета под руководством Родмонги Кондратьевны Потаповой.

Литература

Женило 1995 – Женило В.Р. Компьютерная фоноскопия. М.: 1995.

Женило 2006 – Женило В.Р. Информация, звук и преобразование Фурье-Гаусса // Сб. трудов XV Межд. конф. «Информатизация и информационная безопасность правоохранительных органов». М.: 2006.

Информатика 1994 - Информатика: Энциклопедический словарь для начинающих / Сост. Поспелов Д.А. М., 1994.

МЭС 1998 - Математический энциклопедический словарь. М., 1988.

Потапова 2006 –Потапова Р.К., Потапов В.В.. Язык, речь, личность. М., 2006.

СИС 1988 - Словарь иностранных слов, М., 1988.
Чуи 2001 – Чуи К. Введение в вейвлеты. М., 2001.