

Психофизическая модель восприятия человека и метрика относительного искажения информации.

Пименов И. В.

Как известно, ни сжатие, ни просто передача изображения без психовизуальной модели не возможна. Например, в кодировании видеосигнала была сделана большая работа по психовизуальному представлению цвета, и были разработаны форматы RGB, YUV, CMYK... То же самое для телевидения PAL, SECAM, NTSC. В частности именно поэтому YUV лучше, чем RGB, PAL лучше SECAM для телевидения. В области зрения дальше точной модели первичного представления без потери качества дело не зашло. В области аудио восприятия психоакустическая модель установлена весьма полно и используется в стандартах сжатия с потерями, но базиса для понимания, почему она такая, а не другая, нет. Данная статья предлагает ответы на поставленные вопросы.

Основной постулат данной психофизической модели восприятия человека заключается в тезисе: "**Человек воспринимает информацию**". Для того, чтобы продолжить дальше, нужно определить сущность данных терминов.

Восприятие человека.

Воздействие на рецепторы человека вызывает *раздражение* нейронных окончаний. В свою очередь *раздражение* создает у человека *субъективный эффект - ощущение*. При исследовании восприятия изменений силы раздражения было установлено, что *одинаковые относительные изменения интенсивности раздражения I вызывают одинаковые абсолютные изменения ощущения*, то есть ощущение E пропорционально логарифму раздражающей силы

$$E = k \log I + C.$$

На пороге восприятия $I=I_0$, ощущение равно нулю, поэтому

$$C = -k \log I_0.$$

Или всю формулу можно записать как относительное выражение

$$E = k \log I/I_0.$$

Впечатление - это раздражение высшей коры головного мозга, возникающие при сравнении одного ощущения с другим. Впечатление возникает при сравнения нового ощущения с предыдущими. Впечатления описываются словами относительно предыдущего опыта человека. Можно выделить несколько классов впечатлений. Самыми интересными являются *относительные впечатления*, которые возникают при сравнении тестового раздражителя с тестируемым раздражителем, полученным после искажения первого. Раздражение может нести *информацию*, а может не нести информации. Искажения разбиваются на два класса. *Существенные искажения* воспринимаются человеком как раздражитель, несущественные не воспринимаются как раздражение. Существенные искажения подразделяются на артефакты, смазанности и шумы. Артефакты приводят к возникновению ложного ощущения возникновения информативного раздражения там где его нет. *Артефакт* - это ложный новый объект, приводящий к возникновению ложной информации. Смазанности приводят к исчезновению ощущения по причине исчезновения информативного раздражителя. *Смазанность* - это исчезновение объекта, приводящие к исчезновению реальной информации. Шумы ощущаются как ложный неинформативный раздражитель. *Шум* не приводит к изменению количества объектов.

Изменение количества информации I , в случае искажений, можно записать в следующем виде

$$\Delta I_{\text{смазанность}} < \Delta I_{\text{шум}} < \Delta I_{\text{артефакт}},$$

при этом изменения будут отрицательными для смазанностей, и положительными для артефактов.

Ощущения человека существенно зависят от *области восприятия*. Вследствие того, что *связи между нейронами имеют низкую пропускную способность, возникает необходимость в адаптации разрешающей способности*, достигаемое за счет изменения

чувствительности нейронов. Такое явление проявляется как *концентрация внимания* на области восприятия. Область восприятия достаточно мала, но, для зрения, человек ощущает что он видит достаточно большую область из-за спонтанных изменений области восприятия, вызываемых движениями глаз, примерно два раза в секунду, называемые *саккадами*. Область восприятия является следствием ограниченности физической способности по переработке информации человеком. На этом эффекте основано явление *маскирования и демаскирования*. В области восприятия более сильный раздражитель, который несет больше информации, обращает на себя внимание сильнее, чем более слабый, маскируя его и делая не заметнее. Соответственно, более сильный раздражитель демаскируется более слабым, даже если более сильный является паразитным. Маскирование переводит достаточно слабые существенные искажения в несущественные, а демаскирование переводит достаточно сильные несущественные в существенные. Это явление наблюдается как в пространственной области восприятия, так и во временной.

Ясно, что *информация определяется событиями* в некотором пространстве событий. *Пространство событий* - это *информационное пространство*, которое действует на раздражитель в четырехмерном пространстве-времени и оно *определяется своей статистикой*. Для его описания нужно ввести некоторые определения.

Определения из теории вероятностей.

Статистическое по выборке пространство определяется, как пространство, чьи статистические характеристики полностью описываются выборкой и не меняются в этой выборке. То есть все моменты совпадают со средними по выборке. *Среднее по выборке N значений из области M пространства* определяется как функционал над $y=y(x)$

$$\langle f(y(x)) \rangle = \frac{\sum f(y(x))}{N} = \frac{\int f(y(x)) dx}{\int dx},$$

или более обще через отношение к функции $p(x)$

$$\langle f(y) | p(y) \rangle = \frac{\sum f(y)p(y)}{\sum p(y)} = \frac{\int f(y)p(y) dy}{\int p(y) dy},$$

где интеграл всегда понимается в смысле Лебега по всем значениям аргумента.

Вероятность $p(x)$ может определяться разными путями.

Математическое ожидание - это начальный момент первого порядка, определяемый как среднее значение

$$\mathbf{M}[p(y)] = \int y p(y) dy = \langle y | p(y) \rangle = \langle y(x) \rangle = \mu.$$

Мощностью называется начальный момент второго порядка

$$\mathbf{P}[p(y)] = \int y^2 p(y) dy = \langle y^2 | p(y) \rangle = \langle y(x)^2 \rangle = \mu.$$

Локальное отклонение есть разница между текущим значением и его математическим ожиданием (центром)

$$\mathbf{L}[p(y)] = y - \mathbf{M}[p(y)] = y - \mu = \lambda.$$

Его квадрат представляет из себя *локальную мощность отклонения* λ^2 . Среднее локальное отклонение от центра есть среднее отклонение, и оно совпадает с центральным моментом первого порядка для стационарных пространств.

Соответственно, *средняя мощность отклонения* или центральный момент второго порядка называется *дисперсией*

$$\mathbf{D}[p(y)] = \int (y - \mathbf{M}[p(y)])^2 p(y) dy = \langle \lambda^2 \rangle = \sigma^2.$$

Корень из этой величины представляет из себя *стандартное отклонение* σ . Нормализация локального отклонения стандартным отклонением λ/σ дает *нормальное отклонение* v .

Ковариация представляет взаимный центральный момент

$$\mathbf{Cov}[p_1(y), p_2(z)] = \int \lambda_y \lambda_z dy = \sigma_{yz}.$$

Коэффициент корреляции есть нормированная ковариация и представляется как взаимный нормальный момент

$$\mathbf{R}[p_1(y), p_2(z)] = \sigma_{yz}/\sigma_y \sigma_z = \int v_y v_z dy = \rho.$$

Условная вероятность — вероятность одного события x при условии, что другое событие

у уже произошло

$$p(x|y) = \frac{p(x * y)}{p(y)},$$

где $p(x*y)$ *совместная вероятность* двух событий.

Вероятность не обязательно может быть заданна по одному пространству, но также может быть заданна по нескольким пространствам. Тогда *относительная вероятность* $p(x;y)$ будет взаимной условной вероятностью, причем вероятность p определяется по пространству X

$$p(x;y) = p_x(x|y)p_x(y|x) = \frac{p_x(x*y)p_x(y*x)}{p_x(y)p_x(x)},$$

где все события принадлежат разным пространствам $x \in X$, $y \in Y$. Последнее определение показывает, что по своей структуре относительная вероятность является совместной вероятностью.

Количество информации.

Из наблюдений установлено что *частые, ожидаемые события несут мало информации, и, наоборот, редкие, неожиданные события, обладают высоким информационным содержанием.*

Отсутствие информации определяется как ожидаемое событие с высокой степенью вероятности. И обратно, наличие информации есть неожиданность события. Чем больше информации несет событие, тем более оно невероятно.

Количество информации источника определяется распределением вероятностей своих событий. Количество информации события можно представить как невероятность этого события. Например, по формуле вероятности неосуществления данного события

$$I_p[p(x)] = 1 - p(x).$$

Данное определение иногда используется как мера количества информации. Вообще же *общая формула количества информации определяется из условий, что она должна быть непрерывной, неотрицательной, возрастать при росте невероятности.* Для любой непрерывной возрастающей функции ϕ она имеет вид

$$I_\phi[p(x)] = \phi(1) - \phi(p(x)).$$

Если дополнительно поставлено условие что *информация двух независимых событий с вероятностями p_1 и p_2 и совместной вероятностью $p_1 * p_2$ должна быть равна сумме их информаций*, то *единственной функцией удовлетворяющей данному условию является логарифмическая функция.* Которую можно получить также из соотношения раздражения и ощущения установленной нейрофизиологами, и тут мы тоже приходим к выводу, что для ϕ лучше всего подходит логарифмическая функция, которая связывает восприятие человека и информацию

$$I[p(x)] = -\log p(x) = \log 1/p(x).$$

Это наиболее распространенное определение количества информации. Она также совпадает с определением частной энтропии введенной Шенноном. Вообще же *энтропия* это мера неопределенности источника выражающиеся через среднее количество информации, приходящиеся на один символ источника.

$$H[p(x)] = \langle I[p(x)] | p(x) \rangle = \dot{H}.$$

При каком основании логарифма будет наибольшая информация при наименьшей энтропии? Это задача о наиболее плотной упаковке чисел в системе счисления.

Решается она следующим образом, если взять n знаков, а за основание принять некоторое число x , то получится $m = n/x$ разрядов, и максимальное количество чисел, которые при этом можно записать, будет равно

$$y(x) = x^m = x^{n/x}.$$

Что можно переформулировать как: найти сторону x прямоугольника с площадью $n = mx$ при котором функция $y(x)$ максимальна.

Тогда, для нахождения экстремума нужно приравнять нулю производную от $y(x)$

$$\frac{dy(x)}{dx} = \frac{-n}{x^2} x^{n/x} \ln x + \frac{n}{x} x^{n/x-1} = nx^{n/x-2} (1 - \ln x) = 0.$$

Поэтому для основания логарифма лучше всего взять e . В таком случае эта система счисления называется *натуральной*, а разряд системы счисления называется *Натом*.

Определение количества информации данное выше иногда называют *абсолютной информацией*, поскольку вероятность в нем не соотносится с другим источником, если имеется информация измеримая относительно другого источника, то такая информация называется *относительной*, и вводится формулой

$$I_{\varphi}[p_1(x), p_2(x)] = \left| \varphi(1) - \varphi\left(\frac{p_1(x)}{p_2(x)}\right) \right|.$$

Как мы помним, сила ощущения является относительным логарифмом, поэтому далее в качестве φ всегда используется натуральный логарифм.

Определим величину относительного искажения исходя из вышесказанного.

Величина относительного искажения $D(X, Y)$ между двумя пространствами X, Y есть средняя относительная информация по относительной вероятности искажения каждого пространства относительно другого.

Для того чтоб иметь максимальную чувствительность к информации нужно воспринимать ее в наиболее тяжелом случае. Что соответствует максимальной неопределенности математического ожидания информации, то есть максимальной энтропии $H[p(x)] = \max$. Продифференцировав по x приходим к следующему уравнению $H'[p(x)] = 0$ из которого получается соотношение $p(x) = \langle x | p(x) \rangle$, решая которое мы получаем *формулу Гаусса*.

Для описания пространств условимся, что значения точек пространства лежит в конечном диапазоне значений $0 \leq x < \text{MAX}$. *Одномерное пространство с нормальным распределением* описывается следующей формулой Гаусса,

$$\varphi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}.$$

Его энтропия будет

$$H_{\varphi} = \ln(2\pi\sigma^2 e)^{1/2}$$

и эта энтропия максимальна.

Поскольку информация есть величина не отрицательная, то логарифм от σ^2 не должен быть меньше нуля, то есть σ^2 не должен быть меньше некоторого порога, который может быть равным единице или $(2\pi)^{-1}$. В природе это реализуется всегда из-за того, что область восприятия бесконечна, когда же область восприятия конечна у нейронов, то из-за естественного уровня шума нейронов.

Относительная вероятность искажения так же является степенным нормальным распределением, причем, из-за условия неизменности вероятности в выборке $p_x(a) = p_x(a^*a)$, и определяется как

$$\varphi(x; y) = \frac{\varphi(x)^{xy} \varphi(x)^{yx}}{\varphi(x)^{x^2} \varphi(x)^{y^2}} = \frac{\varphi(x)^{2xy}}{\varphi(x)^{x^2+y^2}} = \varphi(x)^{-(x-y)^2},$$

где y находится в том же диапазоне что и x . Из этого результата виден смысл относительной вероятности искажения, он показывает вероятность возникновения искажения в точке x при некотором отклонении. Чем больше отклонение, тем выше вероятность восприятия данного искажения. При совпадении x и y вероятность возникновения искажения минимальна, но не равна нулю.

Зрительное восприятие.

Рассмотрим зрительное восприятие и явления маскирования и демаскирования возникающие при восприятии видеоинформации.

Для начала рассмотрим пространственное восприятие. Из соображений, что психофизическое пространство по амплитуде имеет Гауссову статистику (статистическое по выборке информационное пространство с нормальным распределением событий), мы получаем, что наименее вероятно сильное изменение амплитуды соответствующее появлению границы. То есть границы несут наибольшее количество информации. Отсюда сразу следует, что граница маскирует мелкие искажения и чем сильнее разница амплитуды на границе и больше количество границ в области восприятия, тем сильнее маскируются искажения и тем незаметнее некоррелированные искажения. Демаскирование наоборот проявляется там, где отсутствуют границы, а небольшое искажение приводит к ощущению возникновения границ, как бы появляются не существующие детали на изображении. Более того, если около границы нет искажений, то возникает ощущение что граница резче (контрастнее) чем на самом деле, тут действуют одновременно эффекты маскирования и демаскирования. По исследованиям нейрофизиологи человек действительно не видит внутренний фон, а видит только границы его, и восстанавливает внутренности по данным границ. Причем исчезновение внутренностей происходит еще в глазу.

Во временной области восприятия появление информации связано с движением некоторого объекта по фону или со сменой сцены, но не с окружающей средой. Первое связано с непрерывностью изменения гауссовой статистики во времени, второе, если статистика меняется скачком. Исчезновение информации связано с восстановлением фона, либо с окружающей средой. Если в окружающем пространстве мало информации, то любое изменение несет больше информации, чем при большом количестве информации в окружающем пространстве. То есть главным носителем информации во временной области восприятия является движение, и человек видит именно его. Если нет движения, то человек не видит ничего, при этом он даже не ощущает, что он чего-то не видит. Для него информация отсутствует полностью, в чем легко убедиться, если вспомнить, что в глазу человека находится слепое пятно, которое человек не замечает полностью. Так как же человек видит, если объект не двигается? Например, лягушки не видят, если нет движения. Интересен сам механизм этого восприятия. Он основан на особенности нашего глаза совершать спонтанные микродвижения, совершаемые несколько раз в секунду, называемые *микросаккадами*. В чем легко убедиться, если закрыть глаз контактной линзой с изображением. Через секунду-две изображение исчезает, и человек просто ничего не видит. Таким образом, природа экономно решила проблему восприятия отсутствия движения и передачи только информации.

Но не только глаз воспринимает информацию, но и высшая кора тоже оперирует информацией, правда более инертно, чем глаз. На этом явлении основан эффект замыливания восприятия: если в течении долгого времени смотреть на одну не меняющуюся картинку, то через некоторое время человек перестает ее воспринимать и при возникновении нового объекта или движения он не замечает его, высшая кора просто не успевает среагировать. К тому же человек не может вспомнить, когда появился тот или иной объект. Это явление иногда называют эффектом часового или эффектом последствия.

Если информация одного объекта равна информации другого, то есть $I[a] = I[b]$, то данные объекты не различимы, например если одна картинка ярче другой на некоторую величину, то человек не может увидеть эту разницу. Вообще человек видит относительную яркость на границах картинок и только благодаря им он может распознать различие в яркости, это явление называется *относительной контрастной чувствительностью*. На этом свойстве основаны разные парадоксы, когда на одной части картинки цвет кажется темным, а на другой тот же цвет светлым. А также если вспомнить, что экран телевизора в выключенном состоянии может быть белым, но во включенном состоянии мы видим на нем черные насыщенные цвета. В описанном случае, несмотря на то, что традиционные метрики искажения данных дадут здесь большую величину, информация же была искажена минимально и только благодаря границам может быть воспринята.

На явлении того, что человек воспринимает информацию с опорой на границы в пространстве-времени, возникают многие парадоксы восприятия. Например, парадокс

монолитного заполнения, такой, что когда смотришь на большую регулярную картинку, заполненную черно-белыми линиями или квадратами (типа дом ...), то возникают скачки картинки в глазу (рябит в глазах, глаза сводит, может возникнуть головокружение, и только перевод зрения может восстановить восприятие). Это ограничение нашего восприятия - мозг перестает видеть внутренности, а границы далеко или вообще не воспринимаются. В этом парадоксе и в иллюзиях движения теряется ощущение пространства, человек видит попеременно все равновозможные картинки с одинаковой информацией.

Человек не ощущает больше чем $e^e \approx 16$ градаций интенсивности в области восприятия данной точки. Это связано с наиболее плотной упаковкой информации. Ощущения большого количества градаций возникает из-за меняющейся во времени и пространстве области восприятия, которое называется *адаптацией*. Благодаря адаптации человек нормально воспринимает приблизительно $e^e \approx 10^6$ градаций интенсивности в фототопическом зрении, а со скототопическим до 10^{10} . Плотность нейронов не равномерна в сетчатке глаза. Она максимальна в районе центральной ямки и минимальна на краю сетчатки. Поэтому, наибольшее ощущение градаций интенсивности наблюдается в центре зрения, где шаг воспринимаемый шаг градаций интенсивности минимален, и наименьшее ощущение градаций интенсивности видно на периферии зрения, где воспринимаемый шаг градаций интенсивности максимален. Отсюда наблюдается эффект, что слабая вспышка света обращает на себя внимание на периферии больше, чем в середине.

Из-за разницы в плотности нейронов в сетчатке, человек обращает свое внимание не равномерно. Одни события могут быть видны лучше, если человек смотрит под одним углом, а другие под другим. К тому же, чем больше концентрация внимания на одном участке, тем меньше воспринимаются другие. Чем выше концентрация внимания, тем выше чувствительность к информации в области восприятия. Если человек концентрирует свое внимание в высокой степени на маленьком участке или событии, то он перестает видеть изображение, и может различить даже отдельные точки в этом участке. Поэтому предполагать, что человек на всем концентрируется одинаково максимально не возможно. Например, был проведен опыт, в котором людей заставляли следить за мячом, а в центре поместили человека переодетого обезьяной. Все испытуемые правильно подсчитали количество пассивов, но никто не заметил обезьяны, хотя она занимала пол экрана. Более общее правило: чем выше концентрация внимания и/или меньше область восприятия, тем меньше воспринимается изображение и больше отдельные точки, события. Из этого следует известный факт, по которому, все парадоксы-иллюзии могут разрушаться при выборе подходящего уровня концентрации внимания и области восприятия. Причем, чем первичнее эффект, то есть ближе к сетчатке глаза возникает явление, тем сложнее его разрушить, для этого надо больше концентрации внимания и меньше область восприятия. И обратно, чем возникновение эффекта ближе к высшей мыслительной деятельности, то есть что возникает вследствие привычки (такой как трактовка размеров изображений, распознавание лиц), тем он легче разрушается.

Все выше сказанное относится также к аудио восприятию. Только вместо двухмерного пространственного восприятия, где частота света воспринимается четырьмя (YRGB) типами клеток одновременно с временными зависимостями, звуковое восприятие одномерно по частоте (высоте) звука с одним типом клеток и по времени с другим типом клеток. В остальном, все тоже самое.

Формула количества информации для психофизического пространства получается, подставив формулу Гаусса в формулу количества относительной информации по отношению к полностью пустому пространству. Поскольку количество информации должно измеряться в самом сложном случае, то количество информации будем выражать в натуральной системе счисления с использованием натурального логарифма, что дает формулу

$$I[\varphi(a), \varphi(0)] = \left| -\ln \frac{\varphi(a)}{(2\pi)^{-1/2}} \right| = \frac{1}{2} \left(\frac{(a - \mu_a)^2}{\sigma_a^2} + \ln \sigma_a^2 \right),$$

где μ_a , σ_a^2 вычисляются в области восприятия точки а. Из нее видно как можно для зрения проинтерпретировать моменты входящие в формулу. Тогда математическое ожидание μ_a является локальной яркостью в области восприятия точки а, а стандартное отклонение σ_a является локальной контрастностью в области восприятия точки а. Как видно информация передается двумя способами, нормальным отклонением ν_a и стандартным отклонением σ_a . Это точно отражает тот факт, что тон может быть представлен как относительная амплитуда (смещение красок, фонов) и как относительная частота (смещение точек, фаз), что мы видим в телевидении и в печатной продукции. Также видно, что яркость не воспринимается локально, а только за счет своих границ, тогда как контрастность воспринимается локально, но значительно слабее, чем край. Как легко проверить, из этой формулы также следуют все остальные явления связанные с маскированием и демаскированием восприятия в подходяще определенной области восприятия пространства-времени.

В дальнейшем информация поступает на следующие уровни восприятия, которые также могут оперировать относительными информациями, в частности, если данная информация соотносится с информацией о дисперсии или с большей областью восприятия, то формула относительной информации получается близка к Лапласиану Гауссиана, соответственно возникают области возбуждения и торможения, что, как известно, хорошо описывает зрение.

Нейронный механизм возникновения области возбуждения и торможения основан на том факте, что формула относительной информации содержит взятие абсолютного числа. Эта операция разлагается на сумму максимумов в виде

$$\mathbf{I} = |\mathbf{I}_{+-}| = \max(\mathbf{I}_{+-}, 0) + \max(-\mathbf{I}_{+-}, 0),$$

где каждый максимум представляет из себя нейрон. А \mathbf{I}_{+-} выражение относительной информации со знаком относительно одной точки, но с разными областями восприятия

$$\mathbf{I}_{+-}[\varphi_-(a), \varphi_+(a)] = -\ln \frac{\varphi_-(a)}{\varphi_+(a)} = \frac{1}{2} \left(\frac{(a - \mu_{a+})^2}{\sigma_{a+}^2} + \ln \sigma_{a+}^2 - \frac{(a - \mu_{a-})^2}{\sigma_{a-}^2} - \ln \sigma_{a-}^2 \right),$$

где знаком плюс помечена большая область восприятия, а минус – меньшая. Отсюда сразу видно как нейрон сжимает информацию. Если стандартные отклонения $\sigma_{a-} = \sigma_{a+}$ и/или нормальные отклонения $\nu_{a-} = \nu_{a+}$ меньшей и большей области восприятия равны, то информация об этом не передается. Тем самым достигается та самая низкая пропускная способность канала зрения, что так поражает нейрофизиологов.

Метрика относительного искажения информации.

Найдем метрику относительного искажения информации между пространствами А и В, статистика которых должна описываться *нормальным распределением* (Гаусса). Для этого подставим формулу Гаусса с относительной вероятностью искажения в формулу относительного количества информации.

$$D(a,b) = \mathbf{I}[\varphi(a;b), \varphi(b;a)] = |-\ln \frac{\varphi(a;b)}{\varphi(b;a)}| = \left| \frac{1}{2} \left(\frac{(a - \mu_a)^2}{\sigma_a^2} - \frac{(b - \mu_b)^2}{\sigma_b^2} + \ln \frac{\sigma_a^2}{\sigma_b^2} \right) (a - b)^2 \right|,$$

где точки $a \in A$, $b \in B$. Моменты вычисляются в области восприятия (некой окрестности) точек указанных внизу в индексах.

Отсюда, если пространства совпадают, то $D(a,a) = 0$.

Это и есть формула для одной точки пространства. Формула для искажения всего пространства есть среднее по всем точкам.

Анализ получившегося результата.

Равенство $D(a,b) = 0$ называется порогом несущественных искажений. Если, искажение равно порогу, то человек не воспримет его как артефакт. Что важно для

преобразований медиаданных и сжатия с потерями. Если искажения информации лежат на пороге, то сжатие максимально при отсутствии потери качества (*минимакс потеря-качество*). Это достигается если локальная мощность отклонения $\lambda_a^2 = \lambda_b^2$ и дисперсия $\sigma_a^2 = \sigma_b^2$ равны. Причем знак отклонения не важен, поскольку только границы несут информацию. Это обычное явление при сравнении аудиоданных, где имеется частотное кодирование. В этом случае мощности отклонения могут быть равны, а информация передается дисперсией.

Данную формулу можно упростить двумя путями. Первый путь заключается в том, что верхняя формула симметрична по обоим пространствам, поскольку пространства взаимно относительно, но можно ввести несимметричность по пространствам и убрать взаимность информации. Тогда искажение будет мериться как отношение информации об относительном искажении равномерно шумящего пространства к информации об эталоне, и формула принимает вид

$$D_0(a,b) = \mathbf{I}[e(a;b)] \mathbf{I}[\varphi(a)]^{-1} = \frac{-\ln e^{-(a-b)^2}}{-\ln \varphi(a)} = \frac{(a-b)^2}{\frac{1}{2} \left(\frac{(a-\mu_a)^2}{\sigma_a^2} + \ln 2\pi\sigma_a^2 \right)},$$

где a нужно рассматривать как эталонную точку, а b как тестовую точку. В этом случае информация только с эталонного пространства имеет значение. В этой формуле не сравниваются информации точек, а только их относительные отклонения относительно эталона, поэтому, пропадают эффекты маскирования по совпадению информации и минимакс потеря-качество. Но данная формула позволяет оценить чистую величину искажения от эталонного пространства, даже не зная информацию тестовой точки.

Другой путь заключается в отказе от относительных вероятностей. Производится только сравнение информации точек без непосредственного их сравнений, по формуле

$$D_1(a,b) = \mathbf{I}[\varphi(a), \varphi(b)] = \left| -\ln \frac{\varphi(a)}{\varphi(b)} \right| = \left| \frac{1}{2} \left(\frac{(a-\mu_a)^2}{\sigma_a^2} - \frac{(b-\mu_b)^2}{\sigma_b^2} + \ln \frac{\sigma_a^2}{\sigma_b^2} \right) \right|,$$

что позволяет говорить о чистом искажении информации без влияния локальной амплитуды изменения на искажение информации. Этот подход позволяет сохранить все эффекты маскирования/демаскирования и проявить чистое искажение информации.

Определение психофизических параметров.

Для полного определения психофизических параметров нужно научиться вычислять моменты, что равносильно определению области восприятия. Это самое трудное для стандартизации в данной модели, как с психофизической стороны, так и по вычислительным затратам. Трудность заключается в том, что область восприятия человека очень адаптивна и разноуровнева. Эта область зависит от органов чувств участвующих в восприятии. Например, у зрения в глазу один нейрон имеет свою ограниченную область восприятия (называемое *рецепторным полем*), другой другую. Все они имеют свою общую область, которая зависит от центральной ямки и хрусталика глаза. А высшая кора преобразует все это в свою область, которая зависит еще от концентрации внимания на объекте и на расстоянии до объекта, и многих других, таких как освещенности и цветовой температуре для зрения, поворота головы для слуха. К тому же точки в области восприятия не равнозначны. Большая вычислительная сложность и плохая формализованность влияния каждого фактора приводит к потребности упрощения модели области восприятия для практических целей.

Определение области восприятия.

Область восприятия - это выборка вокруг рассматриваемой точки, которая связана (не имеет разрывов), компактна (не имеет полостей, не похожа на спираль и тд.), и вообще, выпукла (от каждой точки можно провести прямую проходящую только через точки области).

Вероятность влияния некой точки на данную точку зависит от расстояния. Но нейроны оперируют информацией. Поэтому, можно принять, что информация о влиянии

точки равна расстоянию до нее.

$$I[p(|\mathbf{r}|)] = |\mathbf{r}|,$$

где \mathbf{r} радиус-вектор.

Подставляя в формулу количество информации, и решая данное уравнение относительно $p(|\mathbf{r}|)$, получаем

$$p(|\mathbf{r}|) = \exp(-|\mathbf{r}|).$$

Это и есть выражение для вероятности входящие в моменты для первичного восприятия информации. Подставив данную вероятность в формулы моментов, можно сразу увидеть что, количество информации будет максимальным в области прохождения границ, что хорошо согласуется с данными нейрофизиологии, по которым максимум нейронного отклика достигает при прохождении через него контрастной границы.

Область восприятия зависит от изотропии расстояния и может иметь любую выпуклую форму, чаще всего эллипс. Форма позволяет разным животным еще в глазу выделять типы информации.

Область первичного восприятия у человека это круг (шар) радиус которого тесно связан с метрикой расстояния, зависящий, вообще говоря, от места нейрона. В какой-то момент вес $p(|\mathbf{r}|)$ становится пренебрежимо малым, это и определяет радиус восприятия. Однако, к сожалению, для реальных измерений мы работаем с картинкой полученной в собственной метрике не зависящей от нейронной. Поэтому, мы принимаем эту метрику за единицу. При 8-ми битовой ширине компоненты максимальный радиус восприятия, при котором влияние максимального искажения будет не меньше единицы, получается равным 5-ти. Это хорошо согласуется с тем фактом, что в среднем количество связей нейрона с соседями равняется приблизительно 111. Если же вспомнить, что градаций восприятия одного нейрона на одной точке приблизительно 16, то можно урезать данный радиус до 3-х. На реальных кодерах наиболее часто используется разбиение картинки на прямоугольные области, из-за чего возникают артефакты блочности, то есть появляются горизонтальные и вертикальные линии, к которым человек очень чувствителен благодаря доставшимся от низших животных специальных схем реагирующих на данный тип информации. Чтобы лучше их детектировать и ускорить вычисления, ограничим область восприятия квадратом 5x5. Данный размер области восприятия есть минимально возможный, и соответствует максимальной концентрации внимания в разумных смыслах.

Как указывалось выше, данную метрику можно использовать не только для измерения искажения амплитудного представления тона, но и для фазово-частотного представления, которое передает внутренний тон точками и штрихами, например, в искусственных картинках и в печатной продукции. Что является большим преимуществом. Это является следствием изотропии области восприятия. Тогда внутренний тон, переданный дисперсией, не будет излишне обнаруживаться, а разница будет сильно проявляться на границах области несовпадения представления внутреннего тона. Однако, если вы считаете данный эффект не желательным, то от него легко избавиться, сделав область восприятия не изотропной, тем самым убрав лишние симметрии, например, выбрав область восприятия прямоугольной 5x7.

Также матрицу моментов можно вычислять с учетом временной координаты. Физиологически восприятие времени у человека основано на памяти нейронов, что для практических целей использовать трудно. Поэтому можно рассматривать временную координату как координату пространства. При этом будущее и прошлое не симметричны в восприятии человека. Для оценки текущей точки более важно ее прошлые состояния, чем изменение ее в будущем. Отсюда, в модели, нужно уменьшать радиус восприятия для будущего сильнее, чем для прошлого. Время памяти нейронов составляет примерно 0.1 секунды. Поэтому, можно ввести куб 5x5x4 для области восприятия со временем, но лучше ввести область восприятия в форме призмы 5x5x4, например, с шагом -2, что ближе к человеческому восприятию и позволяет обнаруживать дрожание картинки во времени.

Сравнение IRDM с PSNR.

С помощью программы YUVI измерим крайний левый средний участок, начиная от глаза foremap-qcif (блоки по нулевой колонке начиная со второй строки).



Рис.1. foreman-qcif Y: Слева оригинальное изображение, справа искаженное. Вверху без сетки, внизу с сеткой.

```
yuvi>metric 3 1
IRDM 280 Y:4.021 U:5.439 V:3.767
frame 280 16x16*0x2(22)=Y:5.945 U:6.250 V:4.969 metric=3
frame 280 16x16*0x3(33)=Y:6.234 U:0.563 V:1.656 metric=3
frame 280 16x16*0x4(44)=Y:8.094 U:0.250 V:2.922 metric=3
frame 280 16x16*0x5(55)=Y:8.723 U:4.563 V:3.703 metric=3
yuvi>metric 1 1
PSNR 280 Y:33.244 U:36.814 V:37.817
frame 280 16x16*0x2(22)=Y:31.853 U:37.674 V:37.607 metric=1
frame 280 16x16*0x3(33)=Y:33.645 U:41.752 V:38.192 metric=1
frame 280 16x16*0x4(44)=Y:34.089 U:39.049 V:38.829 metric=1
frame 280 16x16*0x5(55)=Y:31.371 U:35.463 V:33.540 metric=1
```

Как видите, искажения на глазе не сильно заметны (только под бровью видно), в тоже время под глазом на щеке сильно. Однако величина PSNR не отражает этого факта, она говорит что глаз и шея искажены сильнее по амплитуде, чем щека. IRDM же показывает что на глазу искажения будут заметны меньше, чем на щеке.

Итак, найдена адекватная человеческому восприятию формула меры качества. Несмотря на большие вычислительные затраты, чем при использовании таких мер как PSNR, MSE этот метод позволяет адекватно измерить качество. Принципиально не существует адекватного метода измерения качества, который измеряет только по одной точке пространства.

Область восприятия это единственный внешний параметр данной модели определяемый из эмпирико-вычислительных критериев. Что выгодно отличает данную модель от VQM и SSIM метрик, где содержится множество эмпирических и вычислительных параметров. В случае видео, даже PSNR имеет два параметра, амплитуду сигнала и параметр связанный с правилом подсчета метрики во времени.

Данная метрика позволит улучшить сжатие с потерями. Действительно, для того, чтоб улучшить качество при том же количестве бит, необходимо где-то его ухудшить. Если мы увеличим искажение в одном месте и уменьшим в другом, то такие метрики как PSNR покажут, что качество не изменилось. В тоже время, мы можем увеличить искажение в существенном месте и человек увидит артефакт, смазанность или шум, а в другом месте уменьшение искажения человек никак не заметит. PSNR на это никак не отреагирует или отреагирует не адекватно. В то же время IRDM отреагирует адекватно, более того

покажет, где и на сколько можно исказить так чтобы человек ничего не заметил.

Литература:

М. Вернер. Основы кодирования. Москва: Техносфера, 2006.

Р. Гонсалес, Р. Вудс. Цифровая обработка изображений. Москва: Техносфера, 2005.

Ю. А. Ковалгин, Э. И. Вологдин. Цифровое кодирование звуковых сигналов. СПб: КОРОНА-принт, 2004.

Н. Н. Красильников. Теория передачи и восприятия изображений. Москва: Радио и связь, 1986.

Колмогоров А.Н., Гельфанд И.М., Яглом А.М. К общему определению количества информации. - ДАН СССР, 1956, т. 111, с.745.

Колмогоров А.Н. Три подхода к определению понятия "количество информации". Проблемы передачи информации. т. I , вып.1, 1965, с. 3-11.

Прохоров Ю.В., Розанов Ю.А. Теория вероятностей. Основные понятия. Предельные теоремы.

Случайные процессы. Справочник. - М.: Наука.1987.

Д. Хьюбел. Глаз, мозг, зрение.

Д. Марр. Зрение. Информационный подход.

Г. Тамар. Основы сенсорной физиологии.

Ч.А.Измайлов и др. - Психофизиология цветового зрения.

Я.А.Альтман и др. - Слуховая система.

ИЛЛЮЗИИ ЗРИТЕЛЬНОГО ВОСПРИЯТИЯ <http://www.sciam.ru/2004/6/ochevidnoe.shtml>