

ОСНОВЫ КОМПЬЮТЕРНОЙ ОБРАБОТКИ ЗВУКА

Компьютерная обработка звука – новая прикладная область информатики. Алгоритмы обработки звука интересны тем, что результаты их применения можно сразу же «услышать». Среди задач, связанных с этой проблемой, есть как трудные, решение которых доступно лишь профессионалам, так и такие, которые могут успешно решать студенты и школьники.

Редакция обратилась к специалисту в области математических методов обработки звука кандидату физико-математических наук Сергею Витальевичу Рыбину с просьбой рассказать о том, чем занимается эта область информатики. Приводим запись лекции, которая была прочитана в Санкт-Петербургском государственном электротехническом университете для преподавателей математики и информатики средних школ, студентов и школьников.

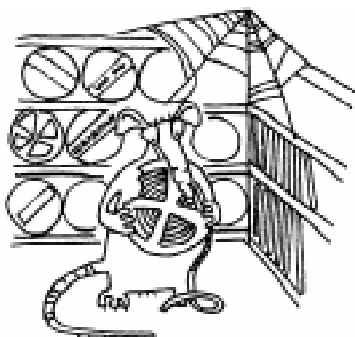
В своем выступлении я постараюсь обойтись без подробной формализации, поскольку формализация в данном случае привела бы к достаточно сложным вещам. Поэтому я буду рассказывать в основном «на пальцах», без привлечения математического аппарата.

Для чего необходима компьютерная обработка звука? Многие из вас, наверное, слышали о существовании архивов звуковых материалов. В Москве, например, есть специальный телерадиофонд, где, начиная с 20-х годов, складировалась та или иная информация. Там, например, есть записи Шаляпина, записи Рахманинова, есть, конечно, и выступления В.И. Ленина, которые тоже представляют историческую ценность. Есть очень много записей духовной музыки. Приводить здесь полный список категорий, хранящихся в фонде, я не буду из-за недостатка места.

К сожалению, аналоговые материалы хранятся чаще всего в неблагоприятных условиях. Вообще говоря, магнитные ленты долго хранить нельзя: происходит процесс саморазмагничивания. Поэтому запись следует перевести в цифры. А если у вас уже

есть цифровой носитель, можно обработать эту информацию на компьютере, что позволит не только сохранить запись, но и существенно улучшить ее качество. Правда, срок хранения на компакт-дисках тоже ограничен, но лет 80 вы можете спокойно хранить информацию, и можно надеяться, что за 80 лет человечество найдет более совершенные способы ее хранения. И даже современная запись, сделанная в студии, никогда не выпускается без обработки: ее подвергают так называемому звуковому ремастерингу (об этом я расскажу чуть позже).

Теперь я хотел бы сказать несколько слов об алгоритмах. Идеология заключается в том, что звукорежиссер сам себе ставит задачу. Звукорежиссер не знает математики, он и в технике разбирается достаточно слабо, но он понимает на слух, чего он хочет. После этого строится математическая модель, но звукорежиссер не может в ней адекватно отразить все, что он хочет. Поэтому все такие алгоритмы строятся математиками с помощью математического моделирования при определенном, иногда очень высоком, уровне эмпирического подбора.



...аналоговые материалы хранятся чаще всего в неблагоприятных условиях...

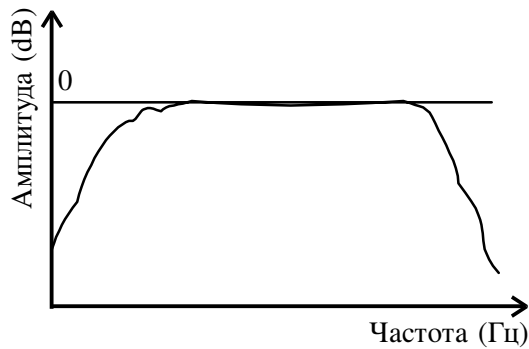


Рисунок 1. Амплитудно-частотная характеристика.

Теперь – об основных понятиях, которыми оперирует наука о цифровой обработке сигналов. Все понятия, связанные с обработкой звука, строятся вокруг понятия спектра. Допустим, вы пытаетесь разложить сигнал в некоторую сумму синусов различной частоты, с различными коэффициентами. Основную роль при этом играет спектр и те понятия, которые прилегают к понятию спектра. Это амплитудно-частотная характеристика и фазово-частотная характеристика. Понятие спектра непосредственно связано с понятием тембра звука. Спектр зависит от частоты, и точки спектра показывают, насколько велик вклад данного синуса в сигнал. Если есть пик, например, на 200 Гц, то эта высота вносит в сигнал основной вклад.

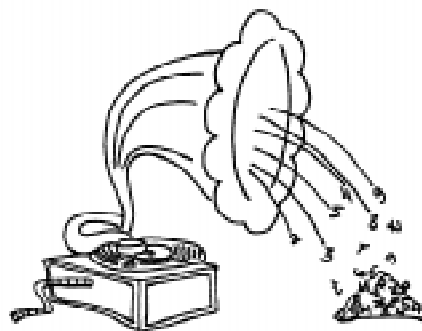
С амплитудно-частотной характеристикой, я думаю, многие знакомы, поскольку на многих приборах изображен соответствующий график. Вопрос о том, насколько он соответствует действительности, я оставлю «за кадром», но суть в том, что этот график показывает «пропускную способность» вашего прибора (рисунок 1).

Например, ноль децибел для некоторой частоты означает, что сигнал на этой частоте передается без искажений. Отрицательное значение показывает, что сигнал уменьшается, искажается на выходе прибора. Амплитудно-частотная характеристика показывает режиссеру, насколько соответствующий прибор искажает входной спектр, что будет со спектром сигнала после обработки на данном уст-

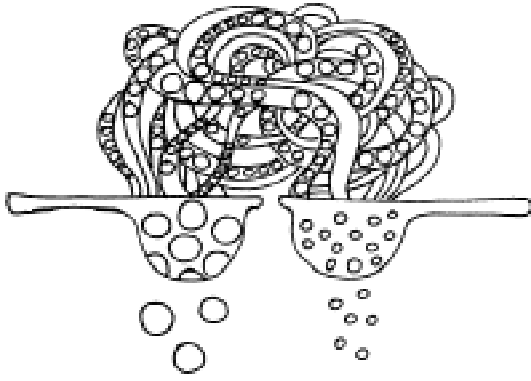
ройстве. Кроме того, для звукорежиссера очень важную роль играет фаза. Если я буду сдвигать некоторый синус, то это соответствует задержке относительно начальной точки. В нашем восприятии фазовые соотношения между стереоканалами формируют так называемую стереопанораму. Изменения фазы человек на графике не видит, но на слух воспринимает очень хорошо. Фаза – это частотная характеристика, показывающая, какой сдвиг по фазе соответствует данному синусу, данной частоте.

Теперь – о том, как перевести звук в цифру, если у нас есть аналоговые носители. Перевод звука в цифру осуществляется с помощью аналого-цифрового преобразователя. Есть аналоговый сигнал (скажем, напряжение), и вы с одинаковым шагом делаете выборку. Получается последовательность точек. Разумеется, реальная схема значительно сложнее.

Встает вопрос о том, с каким шагом делать выборку. Сейчас в мире принята частота 44,1 КГц. Это означает, что за 1 секунду мы снимаем 44100 точек. Есть еще одна базовая частота, менее популярная, это 48 КГц, но 44,1 КГц считается базовой. Почему часто используют именно такую частоту? Это связано с особенностью человеческого уха. Наше ухо воспринимает сигнал от 20 Гц до 20 КГц. Есть теорема Котельникова-Шеннона (она была независимо доказана этими двумя учеными в тридцатых годах), которая утверждает, что для того, чтобы получить спектр без искажений до частоты F , нужно производить выборку с частотой $2F$.



Теперь — о том, как перевести звук в цифру...



Фильтр верхних частот «отрежет» нижние, фильтр нижних «отрежет» верхние...

Поэтому, чтобы обеспечить 20 КГц, нужно взять 40, но берут с запасом – 44,1. Так исторически сложилось, и это связано с особенностями первых устройств. Но даже этого недостаточно. Высокочастотные сигналы, как правило, очень слабы, тем не менее они создают определенный колорит, фон, когда мы воспринимаем музыку или речь, поэтому их тоже нужно записывать без искажений. Теперь представим себе, что такое запись звука на 20 КГц. Мы на каждый его период имеем иногда две, иногда три точки. Поэтому высокочастотные сигналы «плывут». Если сделать выборку с переменной фазой, появляется низкочастотная модуляция. То есть огибающая синуса неровная, она представляет собой низкочастотный синус, и в результате появляется такое легкое «гудение». Чтобы этого не было, сейчас принят новый стандарт – 96 КГц, который вступил в силу в 1996 году.

С этой частотой связано и представление сигнала в компьютере. Сейчас практически все АЦП 16-разрядные. Это означает, что они поддерживают целочисленные сигналы в диапазоне от –32768 до 32767. Есть, в принципе, 24-разрядные АЦП, но они применяются редко. Хотя, на самом деле, и этого недостаточно, поскольку режиссер с хорошим слухом слышит перепад всего в 1 уровень.

Несколько слов о математическом аппарате. Итак, у нас есть запись звука в компьютере, и мы работаем с последова-

тельностью целых чисел в диапазоне от –32768 до 32767. Непрерывный аналоговый сигнал превратился в целочисленную последовательность, в которой он и хранится. Математический аппарат, который используется здесь, довольно простой. Прежде всего, это дискретное преобразование Фурье (ДПФ), которое позволяет получить спектр сигнала:

$$y_j = \sum_{k=0}^{N-1} x_k \omega^{-jk}, \quad \omega = e^{\frac{2\pi i}{N}}$$

Поскольку сигнал дискретный, здесь активно применяются элементы дискретной математики.

Обратите внимание: на входе сигнал был вещественный, а на выходе мы получили комплексный спектр. В этом нет ничего необычного. Дело в том, что значение комплексного спектра учитывает и амплитуду и фазу – отвечает за два параметра. Чем больше N , тем выше разрешение спектра, тем более близко друг к другу лежат эти спектральные компоненты, то есть чем больше мы берем точек, тем ближе можем отслеживать спектральные изменения.

Второй важный момент аппарата – это фильтрация сигнала, так называемый рекурсивный фильтр, который зависит от предыдущих значений (некоторая линейная система, переводящая линейную комбинацию в линейную комбинацию):

$$y_n = x_n + b_1 x_{n-1} + b_2 x_{n-2} + a_1 y_{n-1} + a_2 y_{n-2}$$

Что делает фильтр? Прежде всего – преобразует сигнал. Фильтр верхних частот «отрежет» нижние, фильтр нижних –

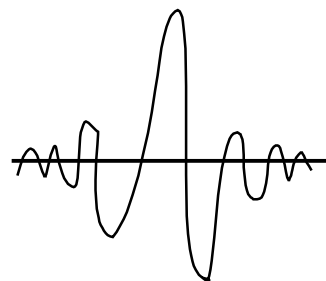


Рисунок 2. Вейвлеты.

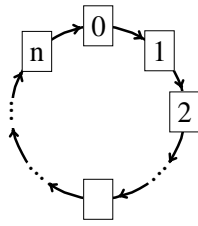


Рисунок 3. Циклическая память.

«отрежет» верхние, полосовой фильтр вырежет полосу, полосно-пропускной эту полосу оставит.

Фильтрация и ДПФ – это, если можно так сказать, классические методы цифровой обработки сигналов. Сейчас появились новые технологии, связанные с применением функций, называемых вейвлетами (от английского слова «Wavelet») (рисунок 2).

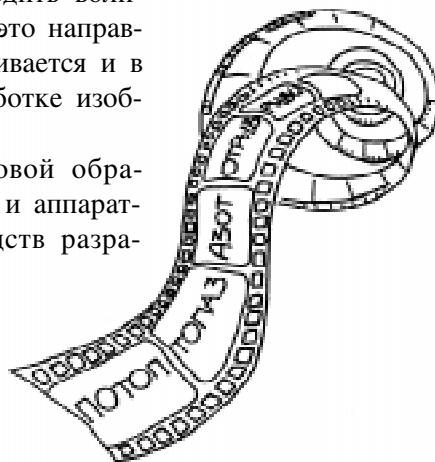
Идея в том, чтобы, вместо разложения сигнала в сумму синусов разложить его в сумму вейвлетов. Это связано с локальностью анализа. Как видите, носитель очень быстро затухает. В чем недостаток преобразования Фурье? В том, что для работы мы должны брать очень большой кадр, и для этого требуется высокое разрешение. Если сигнал будет динамичным, то на кадре будет очень разнородная информация, если преобразование Фурье приходится на участок вблизи музыкального фронта, то получим искаженную информацию о сигнале. А разложение по вейвлетам можно производить вблизи каждой точки. Сейчас это направление очень активно развивается и в обработке звука и в обработке изображения.

С развитием цифровой обработки сигналов появилась и аппаратная поддержка этих средств разработки. До недавнего времени все профессиональные системы, алгоритмы такого рода были разработаны не для общих процессоров, а для так называемых процессоров цифровой обработки сигналов (DSP –

«Digital Signal Processing»). Есть несколько очень известных фирм, которые выпускают DSP процессоры, способные цифровать сигнал и с ним работать. В первую очередь, это Texas Instruments, это Analog Device и известная фирма Motorola. Перечисленные фирмы выпускали и продолжают выпускать специальные средства работы с музыкальными сигналами. Но сейчас появились быстродействующие процессоры Intel, которые позволяют частично отказаться от DSP процессоров, и уже есть ряд программных разработок, которые их не используют. Intel догнал своими возможностями эти спецпроцессоры.

DSP процессоры позволяют вести обработку как целых чисел, так и чисел с плавающей запятой с повышенной точностью. Еще один момент – так называемая циклическая память. Обработка, как я уже сказал, идет по кадрам, и чтобы не потерять информацию о кадре, как правило, работают «со смещением», захватывая соседний кадр. Для этой цели и была разработана быстрая циклическая память (рисунок 3). Помимо этого, DSP процессоры поддерживают много так называемых «параллельных команд», выполняемых одновременно, например, «сложить-умножить». Эффективность такого рода команд можно увидеть на примере формулы ДПФ.

Теперь я перейду к задачам и алгоритмам цифровой обработки звука. Я по возможности охвачу то, с чем сам соприкасался за 16 лет работы со звуком (точнее, с акустическими сигналами). Дело в том, что здесь нет объективных оценок работы алгоритмов. Ни в одной книге по этой теме вы не найдете объективных критериев, поскольку звук – музыка, речь – воспринимается каждым по-своему. До сих пор применяют сравнитель-



...чтобы не потерять информацию о кадре, как правило, работают ..., захватывая соседний кадр...



Сигнал, обработанный прибором, предлагают экспертам...

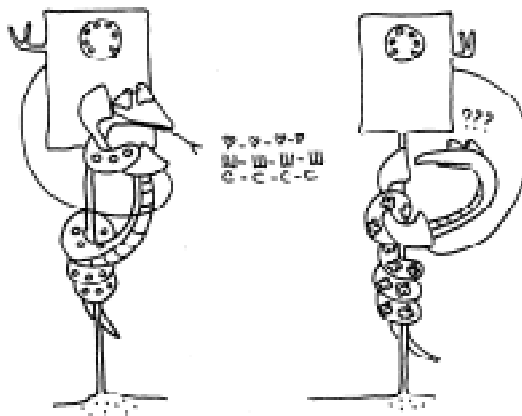
ные (или экспертные) оценки. Сигнал, обработанный прибором, предлагают экспертам. Каждый выставляет оценку, и сигнал, который набирает больше баллов, считается лучшим. Вся оценка алгоритмов обработки музыки и речи чисто субъективная. Один звукорежиссер может сказать, что это полная ерунда, а десять других сказать, что все замечательно. Верят в этом случае десяти.

Какие задачи преследует обработка звука? Я бы разделил их на два класса: задачи речевой обработки и задачи работы с музыкой. Дело в том, что, как это ни обидно для человека, речь устроена гораздо проще, чем музыка, поэтому здесь достигнуты значительные успехи, гораздо большие, чем в музыкальной области.

Первая задача, связанная с обработкой речи, – ее сжатие (для пересылки). Как можно сжимать сигнал? Можно сжимать его по форме, но такие алгоритмы были сразу же отброшены из-за их невысокой эффективности. Есть так называемое частотно-полосное кодирование (это MPEG-формат в музыке). Вы кодируете спектр сигнала, исходя из того, что у вас какие-то частотные полосы более содержательны для человеческого уха, какие-то менее. Если полоса более информативна, вы ее кодируете более качественно. Для человеческого уха, например, известно, что основная полоса – от 300 Гц до 3 КГц. Я приведу один пример. Может быть,

не все знают, что в телефоне стоит фильтр. Спектр передается не полностью. Такое ограничение было сделано для увеличения пропускной способности канала. Тем не менее, вы узнаете по телефону знакомые голоса. Такова особенность человеческого уха: получая такой сигнал, оно восстанавливает его само, проводя как бы синтез, – по обергармоникам, гармоникам высшей частоты оно восстанавливает гармоники основного тона. Это потрясающее свойство нашего уха.

Итак, полосное кодирование учитывает те полосы, которые считаются информационно емкими. Оно сейчас отошло на второй план. В настоящее время для речевого сжатия применяется так называемая вокодерная технология. Здесь используется новый подход к сжатию. Мы не сжимаем спектр, мы не сжимаем форму волны. Я уже говорил, что речь достаточно просто устроена, и есть хорошие модели речи. Они разработаны еще в 60-е годы, а может быть, и раньше (более ранние статьи мне просто неизвестны). Что такое речь? Если я произношу гласную, например, «а», я произвожу комбинацию синусов определенных частот. Она проходит через фильтр, и на выходе я получаю измененный сигнал. Если я произношу согласную, я выдаю некоторый равномерный шум, который тоже проходит через фильтр. Была принята такая идея: не надо кодировать сигнал, надо кодировать параметры модели, то есть коэффициенты



Может быть, не все знают, что в телефоне стоит фильтр...

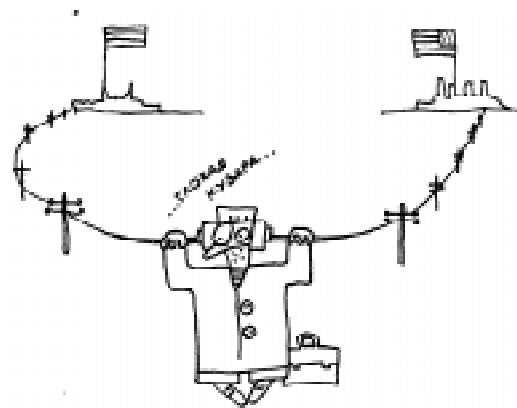
фильтра, которые у каждого человека свои. Я кодирую и передаю коэффициенты фильтра и вместе с ними некоторый основной тон для гласных (тон, с которым я говорю «а»). При этом получается очень сильное сжатие речи. Лучшие программы сжимают примерно в 100 раз. Голос на выходе будет с искажениями, но он все равно будет легко узнаваем. А на этапе синтеза синтезирующая часть принимает коэффициенты фильтра, она знает, что передается – тон, гласная или согласная. Если гласная, то на вход фильтра подается так называемая функция возбуждения, если согласная, – то некоторый шум.

В этой части и заложена проблема, над которой сейчас бьются ученые. Дело в том, что от этой функции зависит голос – насколько он будет близок к истинному? Например, если для всех гласных звуков подавать на вход синтезирующей системы функцию возбуждения, соответствующую одному и тому же звуку, то на выходе получится «робот» (все вы представляете себе «роботную» речь). Недавно мы создали автосигнализацию, которая голосом робота вещает, что происходит с машиной. Дело в том, что возбуждающая функция не дает четкого сигнала, сигнал получается искаженным. Как ни странно, вокодеры хуже звучат на женских голосах. Женский голос более переливчатый, у него очень динамично меняется основной тон. Поэтому все вокодеры проверяются в первую очередь на женских голосах, потому что они считаются самыми сложными.

Вокодерная технология используется сплошь и рядом. Наверняка многие из вас знают про Интернет-телефонию. Некоторое время назад у нас в России был большой Интернет-бум, связанный с Интернет-телефонией, но выяснилось, что у нас очень плохие каналы связи, за исключением нескольких цифровых АТС в городе. Остальные каналы не выдерживают компьютерного звука. Так что пока для России это только фантастика. А реально вокодеры работают в мобильных

трубках GSM. Это один из немногих цифровых мобильных телефонов, большинство мобильных телефонов у нас пока аналоговые. В GSM есть вокодер, который одновременно проводит анализ и синтез: кодирует и передает мой голос, декодирует и синтезирует то, что я получу на выходе. Вокодер в GSM достаточно хорошего качества, потому что сжатие речи не очень большое. Чем больше вы сжимаете речь, чем меньше параметров передаете, тем ниже качество речи, синтезируемой вашим вокодером.

Еще одна задача, над которой сейчас активно работают, – это скремблирование, или, иначе говоря, маскировка речи. Сейчас поняли, что человек, который обладает информацией о другом человеке, обладает очень многим. Поэтому сейчас принято скремблировать разговор. Скремблирование в простейших вариантах появилось еще во время Второй мировой войны при обеспечении связи с руководителями США и других стран. Проблема в том, что скремблирование требует хорошей синхронизации. Когда, например, Путин начинает разговор с Клинтон, предварительно происходит синхронизация двух пунктов – в данном случае, Москвы и Вашингтона. И разговор начинается одновременно – бит в бит. Тогда можно цифровой сигнал рассыпать на биты, перемешать, как угодно, а приемное устройство его обратно перемешает, и получим в точности восстановленный сигнал.



Еще одна задача... – это ... маскировка речи...

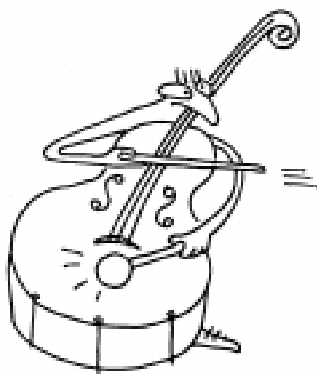
Нужна очень точная синхронизация, которая используется для правительственной связи. Ясно, что обыкновенный человек с телефоном или с трубкой не может себе такого позволить. Поэтому сейчас получили развитие скремблеры без синхронизации. Я начал разговор, нажал кнопку, и с этого момента пошло скремблирование.

Здесь есть простор для творчества, потому что сейчас предложено много алгоритмов скремблирования. Проблема в том, чтобы не было синхронизации. Я предлагаю над этим подумать, потому что защита информации – очень актуальная задача.

Подавление шумов в речи – тоже важное направление. Как правило, разговор идет об узкополосных сигналах, где достигнуты высокие результаты. Например, подавление шума от двигателя автомобиля или самолета во время репортажа. К сожалению, наши отечественные телевизионные станции пока не очень активно этим пользуются. Такие алгоритмы сейчас активно разрабатываются – так называемая адаптивная фильтрация. Вы знаете сигнал, знаете шум и просто из сигнала вычитаете шум. Реально, конечно, дело обстоит сложнее. В этой области достигнуты очень большие успехи (в том числе и специалистами нашего города). К сожалению, в России соответствующие алгоритмы не очень хорошо применяются.

Перейдем к музыке. Обработкой музыки я сейчас и занимаюсь. Я понимаю, что расскажу далеко не все. Какие-то вещи забуду, некоторыми проблемами я не занимался, поэтому мой рассказ ни в коей мере не претендует на полноту.

Первое направление, о котором хотелось бы сказать, – это сжатие музыки. Я уже говорил про формат MPEG. Можно зайти на какой-либо из многочислен-



Любой из нас в состоянии выделить... скрипку на фоне барабана...

ных мультимедийных сайтов, посмотреть, как сигнал разбивается на полосы (кодируется), как устроен алгоритм декодирования (то есть воспроизведение сжатого звука), – все это открыто для любого желающего. (Для того, чтобы узнать адрес, проведите поиск по ключевым словам MP3 или MPEG – и вы получите искомые адреса в большом количестве.) Фирма полностью открыла свой формат. Даже наша команда

в свое время делала попытку сжатия музыкальных сигналов. Этим же занимаются многие солидные фирмы – мне известны работы Delta Telecom, которые обеспечивают сжатие гораздо лучше MPEG. К сожалению, здесь проблема не столько в том, чтобы придумать лучший формат сжатия, а в том, чтобы продвинуть его на рынок, лицензировать. Сейчас MPEG занял свою нишу, и подвинуть новый формат в ближайшее время на этом рынке не представляется возможным, так как это упирается в колоссальные финансовые затруднения. Это пример того, как не самый лучший продукт захватил рынок. Я думаю, всем вам известен другой пример – компания Microsoft, которая обошла всех в плане работы с простыми пользователями, с «чайниками», которых гораздо больше, чем профессионалов. Естественно, все производители программных продуктов должны ориентироваться на любителей, потому что, если бы они ориентировались только на профессионалов, они бы умерли с голоду.

Следующая задача, которая пока не решена и в которой есть простор для творчества, – выделение одного сигнала на фоне другого. Хотелось бы верить, что эта задача разрешима, поскольку человеческое ухо это делает. Любой из нас в состоянии выделить, например, скрипку на фоне барабана, но сделать это программно очень сложно. И хороших, устойчивых алгоритмов решения такой задачи пока не

найден. Задача регулярно ставится и также регулярно не решается: например, оставить голос и убрать оркестр, чтобы наложить на голос другое музыкальное сопровождение. Наше ухо – вообще замечательный инструмент, и если в будущем удастся приблизиться к возможностям уха, это будет просто здорово.

Следующая очень свежая тема, которая сейчас бурно развивается в связи с развитием Интернет, – защита музыки, защита фонограмм. Многие из нас видели, что в Интернете много музыки, много записей. Официально к бесплатному копированию разрешены отрывки до 20 секунд. В течение 20 секунд вы можете использовать любую запись, любой отрывок из фильма, за большее вы обязаны платить. Естественно, о таких прецедентах в России неизвестно – кто же у нас платит? Поэтому сейчас очень бурно развивается так называемая система *water-marker*. На фонограмму ставится бесшумный «водяной знак». Когда вы пропускаете запись через устройство декодирования, выдается уникальный код, по которому можно определить обладателя авторских прав на данную фонограмму.

Защита звуковой информации сейчас очень актуальна, очень бурно развивается, и эта задача по силам студентам. Работа со звуком, мне кажется, интересна студентам тем, что вы можете сразу послушать свой результат. Я начинал работу с акустическим сигналом в интересах серьезных ведомств, и там, пока все дойдет до проекта, уходят месяцы, а то и годы. А «живая» работа со звуком позволяет сразу же оценить свой результат.

И последнее, о чем я хочу подробно поговорить, – это так называемая проблема ремастеринга фонограммы. Это обширная проблема.

Под ремастерингом понимается улучшение, исправление фонограммы и, кроме того, ее монтаж, то есть состыковка фонограммы из кусочков, обработанных разными приборами.

Допустим, вы обрабатываете запись. Например, есть очень много хороших за-

писей на грампластинках, которые до сих пор хранятся в семьях. В Германии недавно стали выпускать специальные приставки к проигрывателям, которые позволяют снимать информацию в цифрах. Итак, вы обработали один фрагмент фонограммы одним прибором, другой – другим, третий – третьим. Что-то взяли из другой фонограммы. И все это вместе называется ремастерингом.

Я уже говорил, что практически любую запись необходимо подвергнуть ремастерингу в студии. С чем это связано? Во-первых, не все записи у нас студийные, очень много концертных записей, которые сразу требуют ремастеринга. В чем причина популярности концертных записей? У нас в стране это не так заметно, но, когда мы работали в интересах западных заказчиков, выяснилось, что там очень много меломанов, которые собирают специфические коллекции, а именно: оркестр А выступил с дирижером В, солистом С в конкретном зале D с программой E. Скорее всего, эти условия никогда больше не повторятся, то есть этот оркестр с этим дирижером, этим солистом, этой программой в этом зале больше никогда не выступит. Это уникальная запись, и таких записей существует колоссальное количество. Достаточно сказать, что в одном Нью-Йорке более 2000 звукозаписывающих студий. В Москве значительно меньше, в Санкт-Петербурге тем более.

А концертная запись гораздо хуже студийной. Я не буду анализировать рок-



...концертная запись гораздо хуже студийной...

концерты, на которых очень много посторонних шумов. Иногда оставляют аплодисменты в паузах, но весь шум, который производят фанаты, приходится убирать. Даже на концерте классической музыки слышны скрип паркета, звонки сотовых телефонов и т.п. (я не говорю об атмосфере зала – она очень важна в концертных записях). Конечно, режиссер сидит на каждом концерте, он использует режиссерский пульт, но после этого все равно надо обрабатывать запись.

Кроме того, все эти алгоритмы ремастеринга оказались востребованы в последние годы, потому что наш громадный кино- и радифонд с началом перестройки начал потихоньку распродаваться. Естественно, было распродано то, что не нужно (или очень легко) реставрировать. Осталось очень много записей очень плохого качества – то, что уже невозможно продать без серьезной обработки. После этого стали бурно развиваться алгоритмы реставрации, шумоподавления.

Какие алгоритмы и методы входят в этот ремастеринг? Основной принцип здесь – «Не навреди»: лучше недоудалить шум, чем вместе с шумом удалить важную информацию и «вместе с водой выплеснуть ребенка». Применяют разные экзотические способы удаления шума. Когда отмечали 100-летие Карузо, разыскали его записи и отреставрировали следующим способом. Разыскали аппаратуру, воссоздали студию начала века и записали там одного певца. Но важен был не певец, а шум тракта, который позволил «вычитанием» удалить шум в записи. Конечно, этот метод дорогой, есть и более дешевые.

Что такое шумоподавление? Это удаление так называемого равномерного шума. Равномерное шипение – это широкополосный шум, который присутствует в каждой области спектра, то есть в каждой частотной составляющей. Еще одна важная деталь – звуковые алгоритмы должны работать в режиме реального времени. Они должны обрабатывать звук, а потом с некоторой задержкой выдавать его

на плейер. Режиссер работает ушами, он не смотрит на пульт или компьютер, он просто двигает рычажок и слушает. Все это можно сделать только в реальном времени. Визуальные приборы лишь помогают звукорежиссеру проконтролировать процесс с технической точки зрения.

Существует два подхода: первый – когда вам известен некоторый шумовой кусочек. Очень часто запись начинается не сразу, и какой-то шум тракта можно послушать. Таким образом, вы получите распределение спектра, получите информацию о спектре шума. Затем, работая по фрейму сигнала, вычитаете шум на каждой частоте.

Вторая ситуация гораздо хуже – когда нет шума. Здесь есть подход, который сейчас часто используется, – подход самообучения. Звукорежиссер рисует шум. Наша задача – создать инструмент для звукорежиссера. Реставрация звука, работа со звуком – удел профессионалов, то есть людей, которых специально этому учат. Как правило, у режиссеров хорошее музыкальное образование и хороший музыкальный слух. Задача инженера, программиста – сделать инструмент в интересах коллектива звукорежиссеров. Конечно, «вычитание» шума – это некоторое упрощение, алгоритм на самом деле более сложный, но это требует более детального его рассмотрения. С «простейшими» алгоритмами можно познакомиться на примере таких программ, как Sound Forge, Dart Pro, Samplitude и других. Их можно найти на большинстве наших мультимедийных дисков.

Следующая задача – удаление так называемых импульсных помех. Например, поцарапали пластинку, у вас идет сигнал, и возникает щелчок. Осыпание магнитного слоя также вызывает помехи. В этом направлении сейчас активно развивается вейвлетная техника, потому что вейвлет – это, образно говоря, и есть импульсный сигнал. Здесь есть три проблемы: нужно обнаружить импульсные сигналы, отличить их от ложных кликов и,

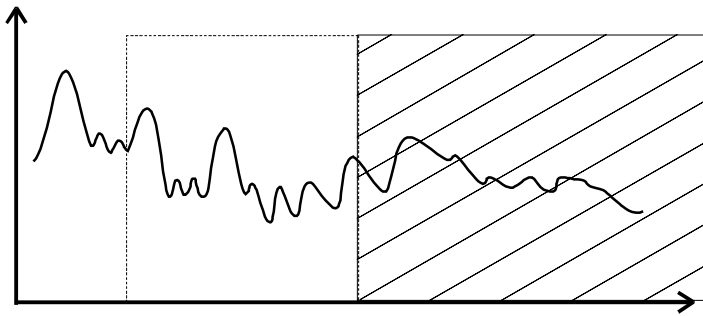


Рисунок 4. Эксайтер.

наконец, чем-то заменить импульсные сигналы. В последнее время часто практикуется подход, когда звукорежиссер на осциллограмме отмечает клики, а затем настраивает прибор, чтобы он сам нашел эти клики или хотя бы соответствующий диапазон параметра. Невозможно прослушать запись в 74 минуты (компакт-диск) – слушать устанешь, поэтому пытаются выделить «характерные» клики и «научить» прибор их распознавать.

Очень интересные приборы эквалайзеры. В бытовых приборах тоже есть простейшие эквалайзеры, в которых спектр разбивается на полосы, и вы можете управлять каждой частью спектра. На бытовой технике больше 6 полос, как правило, не бывает. Недавно мы создали эквалайзер в 4096 полос, то есть с возможностью регулировать отдельно полосы шириной 5 Гц.

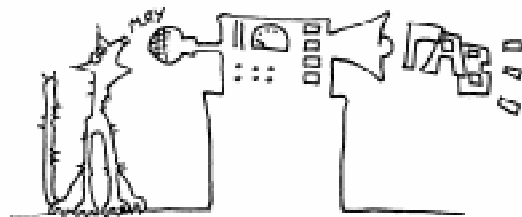
Проблема в обработке музыки заключается еще и в том, что преобразование Фурье линейно, а человеческое ухо воспринимает все не в линейном масштабе, а в логарифмическом. Есть так называемый закон Вебера-Фехнера – это относится не только к восприятию звука, но и к восприятию света, восприятию боли – мы все воспринимаем в логарифмическом масштабе. Например, даже те, у кого нет музыкального слуха, могут отличить «до» первой октавы от «ре». Разница между 440 Гц и $440^{1/2} \sqrt{2}$ Гц очень мала. А, скажем, «до» девятой октавы отличается от «ре» на сотни герц. И чем выше частота, тем меньше наша избира-

тельность, то есть нам не важно высокое разрешение на высоких тонах.

Так что в этом еще один недостаток преобразования Фурье – спектральная шкала равномерна. Вообще-то шкалу нужно логарифмировать, но, к сожалению, до сих пор не придумали быстрых алгоритмов для логарифмической шкалы. Работы в этом направлении

ведутся, и пару лет назад на эту тему на математико-механическом факультете СПбГУ была защищена диссертация, но продвижение в этом направлении очень слабое.

Очень важную роль играет обработка сигнала после шумоподавления. Дело в том, что в процессе реставрации иногда вместе с шумом удаляются и слабые спектральные составляющие полезного сигнала (как правило высокочастотные). Чтобы избежать этого неприятного эффекта, часто прибегают к помощи эксайтера (Exciter). Этот прибор позволяет восстановить неоправданно подавленные высокие частотные составляющие полезного сигнала и повысить качество восстановленной фонограммы. Как же работает этот эксайтер? Звукорежиссер выбирает полосу в спектре, в которой ищутся частоты, обертона которых были «зарезаны» в процессе очистки шума или которых вообще не было в исходной фонограмме. Например, при реставрации записей Шаляпина из-за несовершенства звукозаписывающей аппаратуры тех лет после удаления шума высокие частоты вообще отсутствовали (рисунок 4).



Еще один частотный корректор... предназначен для того, чтобы ... изменить тембр...

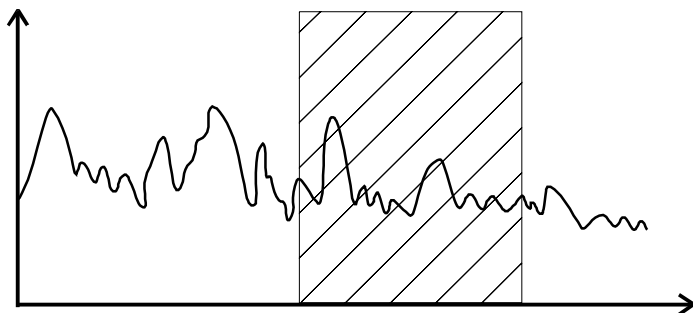


Рисунок 5. Релизер.

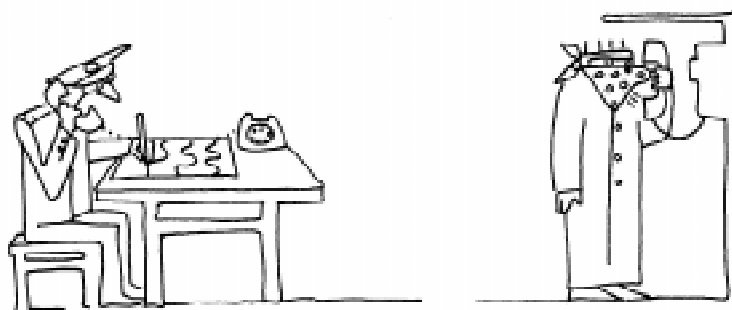
Затем происходит процесс «обогащения» верхней части спектра обертонами исходной фонограммы из выделенной полосы. То есть в данном случае мы пытаемся восстановить истинный голос Шаляпина. Хотя было много нареканий, особенно от любителей старых пластинок, которые говорили, что «Шаляпин так не пел». Но ведь на самом деле никто из них так и не слышал, как поет Шаляпин. Люди привыкли к пластиночному шуму, они считают, что так и должно быть. Когда выпускали диск, была большая дискуссия в прессе о том, что Шаляпин пел не так. А как он пел, – никто уже не знает. Людей, которые слышали самого Шаляпина, уже давно нет в живых. Остались только те, кто слышал его записи на пластинках. И если человек 30 лет слушал пластинку, ему кажется, что все так и было. Здесь, в обработке музыки, психологический эффект играет очень большую роль, и к этому надо быть готовым.

И еще один прибор, релизер (Releaser), часто применяется для восстановления утраченных в процессе ремастеринга гармоник. Известно, что за каждой гармоникой следует ряд ее обертонов (то есть кратных гармоник). При подавлении шума некоторые обертона этого ряда могут «выпадать из ансамбля». Для исправления этой ситуации и применяются релизер (рисунок 5).

В заданной полосе спектра он находит основные гармоники, а затем пытается «вытянуть» их слишком рано угасшие обертона, которые были зарезаны в процессе шумодавления. Релизер также используется для создания звуковых эффектов путем коррекции временных траекторий спектральных максимумов.

И еще две идеи, два алгоритма, которые сейчас очень актуальны. Это time-changing и pitch-transposer. Какова задача time-changing? Есть запись, и вам надо, не изменяя частотных характеристик, чуть-чуть «раздвинуть» запись (замедлить или убыстрить), при этом все тембры должны сохраниться. Если вы просто чуть быстрее или чуть медленнее прокрутите запись, у вас тембр изменится, а этого не должно быть. И такой алгоритм реализован. Для чего это нужно? В свое время мы получили заказ от радиовещательной станции в связи с тем, что у вас часовая передача, вы делаете из нее 59 минут, а 1 минуту пускаете на рекламу. Или, например, на студии по частям записывают классическое произведение. Оркестр сыграл чуть быстрее или чуть медленнее, это надо свести в одну фонограмму. Режиссер вынужден по времени совмещать, чтобы скорость была одинакова.

Еще один частотный корректор, который называется pitch-transposer, предназначен для того, чтобы, сохраняя длительность записи, изменить тембр. Напри-



Вы можете глосовать в трубку, говорить через платок — все равно ваша форматная комбинация останется прежней...

мер, из мужского голоса можно сделать женский, из высокого – низкий и т.д. Я привожу такие гротесковые примеры, потому что на звукостудии этот прибор используется в очень узком диапазоне. Например, в студии записывают инструментальный ансамбль, и концертмейстер по камертону дает настройку «все играют». Иногда он чуть-чуть не так настроил – это будет слышно. И надо изменить тональность этой части. Иногда просто создают звуковые эффекты, наподобие уже упомянутого изменения голоса. Причем человек говорит правильно, с соответствующими речевыми паузами, но у него совершенно другой тембр голоса. Здесь есть возможность увеличения и уменьшения в несколько раз. Такие заказы поступают, в частности, от определенных ведомств, которые всегда интересовались речевыми технологиями.

Я упомяну о еще одной теме, которая сейчас интенсивно развивается. Это так называемый формантный анализ. Дело в том, что у человека есть уникальные характеристики – например, отпечатки пальцев. Еще одна уникальная характеристика человека – набор формантных частот его голоса. Это, грубо говоря, отпечаток человеческого голоса, так называемые звуковые горбы. Есть диапазон, в котором у всех людей поднимается голос. Диапазон этот у всех примерно один, но комбинация горбов – по ширине и высоте – уникальна для каждого человека. Этот набор позволяет однозначно идентифицировать каждого человека, если его голос когда-то был записан (например, голос «телефонного террориста»). Вы можете гнусавить в трубку, говорить через платок – все равно ваша формантная комбинация останется прежней. У нас был такой заказ, к сожалению, не доведенный до конца, когда один цейлонский миллионер хотел сделать распознавание не по пропускам, которые можно украсть, не по

отпечаткам пальцев, а по голосу. Это уникальная система, она, к сожалению, требует создания большой базы данных. Сейчас отдельные такие схемы применяются. В частности, ими обладает МВД.

Я не буду говорить о всех приборах, но скажу несколько слов о ревербераторах. Этот прибор направлен на создание эхоподобных отражений. Если у вас на компьютере хорошая звуковая плата, там есть и встроенный ревербератор.

На самом деле все это очень просто. У вас есть входной сигнал, и, чтобы сделать ревербератор, надо сложить входной сигнал с самим собой, с некоторыми задержками по времени и с некоторыми весами. Я беру этот отсчет сигнала, то есть набор копий сигнала, сдвинутого относительно себя с некоторыми коэффициентами. Этот набор можно измерить для дан-

ного помещения и воспроизвести. Например, мне известен ревербератор, моделирующий акустику Домского собора. Такие измерения можно провести, и после этого, например, воспроизводить свой голос так, как если бы вы выступали в зале с известной акустикой. Это, в принципе, по силам каждому. Можно попробовать разные задержки, получите разные эффекты. Технология очень простая, поэтому в каждой звуковой плате есть встроенный ревербератор, и это как раз то, что можно сделать своими руками.

Есть еще особый вид терапии – лечение звуком и светом, которое часто применяется для лечения, в первую очередь, психически больных. Но конкретные алгоритмы и в этой и в остальных областях, связанных с обработкой звука, обычно не спешат публиковать. Дело здесь не только в стремлении оградить свои разработки от конкурентов, но и в том, что ощущения в восприятии звука в основном эмпирические, поэтому выводы в исследованиях тоже в основном эмпирические.



...в компьютерном варианте можно вырезать ... даже ноты...

И последнее, о чем я хотел бы сегодня поговорить, это не об инструментах, а о монтаже. Если каждый кусочек обработать отдельно, а потом соединить, происходит «щелчок». Для борьбы с таким явлением применяются фейдинги, или склейки. Нужно выделить блок из первого фрагмента, из второго и сложить их с разными весами. Сначала первый, с большим весом, потом веса плавно меняются. То, что сейчас называют фейдингом, долго называли склейками, потому что раньше, до появления компьютерных технологий, звукорежиссер работал с помощью ножниц и специальной пленки для склеивания ленты. Услышит щелчок, на слух найдет это место на ленте, вырежет и склеит концы. Некоторые режиссеры, особенно пенсионного возраста, до сих пор так

и работают – они не доверяют компьютерам. А в компьютерном варианте можно вырезать не только щелчок, но даже ноту. А можно и вставить ноту.

Есть известный анекдот, которым я хочу закончить свое выступление. В студии сидит звукорежиссер и скрипач-неудачник, который пытается записать музыкальное произведение. Режиссер пишет один дубль, другой, третий – ему в конце концов надоедает, и он говорит: «Маэстро, вы сыграйте мне гамму, все остальное я сделаю сам». Сегодня это уже не анекдот – это реалии обработки звука.

Для тех, кто заинтересовался проблемами компьютерной обработки звука, могу посоветовать журнал в Интернете «Звукорежиссер»: http://www.625-net.com/_reg/regmagaudio.htm

Благодарю Дениса Желнина за ценные консультации при подготовке лекции к печати.

НАШИ АВТОРЫ

*Рыбин Сергей Витальевич,
кандидат физ.-мат. наук, ведущий
специалист фирмы «Канонь».*