

НИК СИВЕР

Университет Тафта, США

Тардовская статистика, машинный слух и Big Data. Все, что воспринимаемо, — лишь число

doi: 10.22394/2074-0492-2018-3-193-200

1. Зимой 1857 года немецкий физик Герман Гельмгольц прочитал в Бонне лекцию на тему «Физиологические причины гармонии в музыке». В этой лекции, посвященной связям между музыкой, физикой и строением уха, он дал необычное описание концерта:

Из уст певцов исходят колебания длиной от 6 до 12 футов; у певиц — более короткие колебания длиной от 18 до 36 дюймов. Шуршание шелковых юбок вызывает свои завихрения воздуха, каждый музыкальный инструмент издает свои особые колебания, и все эти системы расходятся сферически, волнами, по направлению от центра, пронизывают друг друга, отражаясь от стен, и в итоге уступают большей силе новых сгенерированных тонов [Helmholtz 1995, p. 57-58].

193

Мир создан вибрациями и резонансами; это, как описал Гельмгольц, «разнородное множество пересекающихся колебательных систем» [ibid., p. 57]. Слух является привилегированным чувством для разделения подобных множества:

Несмотря на то, что это зрелище скрыто от глаза, человек имеет другой орган — ухо, специально адаптированный для улавливания этого процесса. Оно анализирует переплетение колебаний... разделяет несколько тонов, составляющих его, различая голоса мужчин и женщин,

Ник Сивер — PhD, профессор факультета антропологии Школы искусств и наук Университета Тафта. E-mail: Nick.Seaver@tufts.edu

Nick Seaver — professor, Department of Anthropology, Tufts University, PhD in Anthropology, University of California at Irvine. E-mail: Nick.Seaver@tufts.edu
Перевод с английского Марии Матюховой и Антона Смолькина под редакцией Ивана Напреенко и Николая Руденко. Публикуется с разрешения автора. Оригинал текста: неопубликованный драфт выступления на конференции Американской ассоциации антропологов в Сан-Франциско в 2012; доступно по ссылке: <https://static1.squarespace.com/static/55eb004ee4b0518639d59d9b/t/55ece21de4b08279fc0e83ff/1441587741780/seaverAAA2012.pdf>

и даже специфические характеристики тона, производимого каждым инструментом, шуршание платья, звук шагов и т. д. [Ibid., p. 58].

Но ухо не только раскладывает звук на базовые частоты, оно также различает среди голосов и инструментов музыкальный и немusикальный звук. Способность к различению тонов и социальных категорий — гендера и шума — для Гельмгольца соединены в базовой механике уха. Слушание одновременно является биологическим, числовым и социальным — оно резонирует со струнами смычка, математикой синусоидальных волн и колеблющимися объектами окружающего мира.

194

2. Я стою в португальском монастыре и разговариваю с выпускником из Нью-Йорка. Двор покрыт белой металлической крышей и переделан совещательный центр. Мы находимся на международной конференции по поиску и сбору музыкальной информации. Это достаточно разнородное направление, объединяющее IT-исследователей, электротехников, специалистов по библиотечному делу и музыкологии. Выпускник рассказывает мне о своем исследовании: он обучает компьютеры слушать музыку, используя процессор для работы с нейронными сетями. Он уверен, что как только программные мощности станут быстрее и дешевле, использование таких процессоров окупится.

Я с трудом могу его слышать сквозь гул во время перерыва. В машинном слухе — направлении, которое обучает компьютеры слышать — это называется «проблема коктейльной вечеринки». Ее суть заключается в следующем: как можно отделить от остальных голос, который вы хотите услышать? Как правило, человеческое ухо довольно хорошо справляется с этой задачей, а компьютеры пока нет. Я думаю о том, как иначе звучал бы этот дворик, когда был частью монастыря: тихие ежедневные перешептывания вдоль аркад вместо оживленной беседы на ежегодном собрании.

«Музыка является таким же сигналом, как и все остальное», говорит мне выпускник. Аудиофайл в компьютере является длинным списком чисел, который сообщает динамикам, волну какой длины следует воспроизвести. Нейронные сети — и те, что он проигрывает в собранном им самим компьютере в своей лаборатории, и те, что находятся в его голове — используют это список чисел и распознают в нем паттерны. «Музыка — это только разные частоты, взятые в разные временные периоды». Высоты колеблются на одном уровне, ритмы на другом, фразы на третьем. Все эти повторяющиеся структуры скрыты в цифрах потока данных, лишь ожидающих своего математического распознавания. «Математика и музыка!», сказал Гельмгольц. «Самые радикальные противоположности челове-

ской мысли! И все же они поддерживают друг друга! Это все же они демонстрируют скрытый консенсус всех действий в нашем разуме» [Ibid., p. 46-47].

3. Если мы хотим найти философа для решения задач, связанных с вездесущими колебаниями и количественным разнообразием, нам будет тяжело подобрать кандидатуру лучше Габриэля Тарда. В своем труде «Экономическая психология» он описал колеблющийся мир, который резонирует с мирами, описанными мною ранее.

В любом месте присутствуют гармонии, повторяющие себя: колебание на самом деле является мелодичной последовательностью движений, равновесием в движении, отступающих как музыкальная фраза [Tarde, 1969, p. 143].

Хотя основным предметом интереса Тарда были социальные процессы имитации, он понимал их в более широком контексте «универсального повторения» (по выражению самого Тарда). В предисловии к английскому переводу «Законов подражания»¹ Франклин Гиддингс описал интерес Тарда следующим образом:

195

Тард считал, что имитация как социальная норма является единственным видом универсальной деятельности с бесконечным повторением в природе, которую физическом мире мы знаем как волнообразность эфира, вибрации материальных тел, вращение планет на орбите, смена света и тьмы, времен года, жизни и смерти. В этом заключается истина не только социальной науки, но и первый принцип философии вселенной [Tarde, 1903, p. v].

Эти повторения являются фундаментальными для тардовского понимания количественного анализа и науки. Для Тарда мир колебаний представляет собой также мир величин, повторяющихся, и, соответственно, измеримых. Он считал, что без повторения не было бы количественного анализа, аккумуляции похожих единиц для сравнения и, как следствие, науки. Повсеместность вибрации приводит к соприкосновению мира социального и мира природы. Для Тарда мир — это лишь колебания разных масштабов.

Для Тарда связь между социальными и естественными науками наиболее очевидна в сравнении статистики и ощущений:

Почему диаграммы, в которых аккумулярована последовательность преступлений и проступков, прочерченные на бумаге ... должны восприниматься символически, в то время как линия, прочерченная

1 См. также русский перевод — Тард Г. (2011) Законы подражания, М.: Академический проект. — *прим.пер.*

на моей сетчатке полетом ласточки, должна восприниматься как несомненная реальность? [Tarde, 1903, p. 132-133].

Здесь Тард намекает на семиотические последствия его радикального монизма: не существует фундаментальной разницы между символическим движением статистических данных и индексальным движением света на сетчатке. Они различаются по степени, а не по типу. Индексальное не является более реальным или менее произвольным, чем символическое, оно просто быстрее. Статистику трудоемко интерпретировать, к тому же она разносена с описываемым феноменом во времени. Тард предполагал, что в конце концов статистика продолжит «набирать точность, оперативность, объем и регулярность» [Ibid., p. 133] до момента преодоления трудностей, и «статистическое бюро может быть сопоставимо с глазом или ухом» [Ibid., p. 134].

Представляя статистику как науку ощущений, Тард также избражал сенсорное как статистическое:

Каждое из наших чувств своим особым способом дает нам особую точку зрения, статистику внешнего мира. Каждое ощущение — цвет, звук, вкус, и т. д. — является только числом, собранием бесчисленно похожих единиц вибраций [Ibid., p. 134-135].

196

Бруно Латур описывает это как «прогрессивное слияние между технологиями статистических инструментов и физиологией восприятия» [Latour, 2010, p. 156]. Это слияние стало возможным посредством универсальности повторения и, как следствие, количественного анализа. Для Тарда как и для Гельмгольца, а также моего собеседника-выпускника, вычисление является фундаментальным актом восприятия.

Тардеанская статистика дает нам модель для объединенного восприятия нашего знания звука и знания социального, помещая знания акустических практик в социальный контекст, который не является внешним по отношению к вопросам чисел, вибрации или резонанса, но который, напротив, фундаментально встроен в эти вопросы.

4. Алгоритмы машинного слуха основываются на «представлениях свойств» — сжатых версиях аудиоданных, которым придаются аудиальные качества. Если мы будем рассматривать аудиоданные как ряд чисел, показывающих изменения тона говорящего, то представление свойств является резюмированием структур высшего порядка в этом ряду. Резюмирование служит взаимодополнительным целям: она сжимает аудиоданные, делая вычисления более

точными, а также преобразует их в соответствии со стандартами человеческого слуха, в итоге лучше показывая, как воспринимаются музыкальные сигналы.

Стандартное изображение характерных черт в компьютерном слушании — «коэффициенты косинусного преобразования Фурье для частот чистых тонов» (MFCCs). Он был изначально разработан для распознавания речи, изображая 20мс фреймы аудиоданных с набором из 13 чисел, выведенных из серии цепстральных и статистических мэппингов. Джонатан Стерн называет их «техниками восприятия» — технологии, которые интегрируют экспериментальный опыт о человеческом восприятии с вычислительными нуждами технической коммуникации [Sterne 2012]. MFCCs выводит техники восприятия за область слышимого: они представляют характерные звуковые черты в такой сжатой форме, что последние не могут быть воспроизведены в качестве звука — по крайней мере непосредственно.

5. Вернемся в монастырь; я сижу на сессии «Классификация звука». Представленные доклады предлагают различные компьютерные пути для организации аудиофайлов, основываясь на представлении свойств их содержания. Очередной докладчик рассматривает вопрос MFCCs. Из-за того, что они были разработаны для распознавания речи, по мнению докладчика, они пропускают такой музыкально значимый элемент как тон/питч. Разумеется, он предлагает собственную альтернативу.

197

Одной из слабых сторон в тардовском видении сенсорного будущего статистики является то, что он обходит стороной проблемы, которые всегда сопровождают производство чисел. В машинном слухе сочетание вычисления и слушания устроено отнюдь не просто или бесспорно. Повсеместное использование MFCCs сделало их объектом пристального изучения, и презентации в этой сессии предлагают разные альтернативы. Проектировщик показывает таблицу, сравнивающую выполнение этого нового представления свойств в MFCCs. Данная задача предусматривает корректную классификацию большого набора песен. Таблица также демонстрирует, что это новое представление является улучшением в MFCCs, по крайней мере для выполнения этой задачи.

Однако он не останавливается на числовой демонстрации. Как и другие докладчики, он посредством акустической иллюстрации показывает недостатки MFCCs. Я уже упоминал ранее, что MFCCs репрезентации не имеют достаточно данных для воспроизведения напрямую. Для демонстрации, как «звучат» эти наборы, макетные числа MFCCs наполняются и «пересинтезируются» с использованием белого шума.

Он играет отрывок из Кэрол Кинг «You've Got a Friend»¹ через колонки, а затем — MFCC-версию². Ноты на пианино звучат как глухие удары барабанов; голос Кинг едва узнаваем и похож на шипение, «Закрой глаза и думай обо мне, я скоро буду там». Аудитория смеется — это знак дружеской симпатии звучит как угроза ведьмы.

Что делает эту иллюстрацию такой примечательной, так это то, что один из участников указал, что эти аудио-репрезентации не предназначены для воспроизведения, и не существует единого стандарта их пересинтезирования. Только что продемонстрированное воспроизведение по сути не имеет отношения к способности компьютера выделять в нем смысл; этой репрезентации вполне достаточно для обнаружения тона, даже если повторный синтез окажется доказательством обратного. Угрожающее шипение Кэрол Кинг больше связано с акустическими чертами белого шума, чем с MFCCs. Так что тут не происходит формальной критики, скорее неформальная — случайный пример, который скорее иллюстрирует, а не демонстрирует, что один способ представления свойств менее адекватен, чем другой. Почему же ученые упорствуют в аргументах, которые они сами не считают научно обоснованными?

198

Сенсорная статистика Тарда предлагает путь для обнаружения смысла в происходящем. Синтез ощущений и количественного анализа не является односторонним подходом, в котором все феномены сводятся к цифрам. Скорее это обмен между вычислением и чувствованием. Понимая функционирование алгоритмов и уха как аналоговое, исследователи машинного слуха ставят их в один ряд. Научные идеалы, диктующие чистоту методологически созданных чисел, входят в конфликт с аналоговым мышлением, которое связывает в первую очередь слушание с вычислением. Несмотря на преобладание чисел в компьютерном слушании, оценочные способности человеческого уха до сих пор в игре. Слушание — это то, что можно посчитать, но и вычисления — это то, что можно услышать.

Следуя мысли Стефана Хельмрайха [2008], мы можем считать количественный анализ некой трансдукцией — интерфейсом между сферами числового и акустического. Как и другие преобразователи (см., например, [Sterne 2003] о фонографе), количественный анализ зачастую упрощается, натурализуется и принимается как объективный. Однако, как показала моя полевая работа с исследователями машинного слуха, не было обнаружено простой и очевидной связи между вычислением и слушанием. Вернее всего, это спорная сторона перевода, где преобладают количественные/принимаю-

1 <http://nickseaver.net/sound/carolekingWAV.wav>

2 <http://nickseaver.net/sound/carolekingMFCC.wav>

шие опции. Когда исследователи переводят музыку в более общие паттерны, они продолжают спорить, опираясь и на подтверждения со стороны восприятия и со стороны чисел. Внося поправки в философию Тарда, мы можем согласиться, что слушание и вычисления связаны, однако следует признать, что существует множество способов как для вычисления, так и для слушания.

6. «Мы все слушаем музыку исключительно онлайн», утверждает сайт одной из главных музыкальных компаний, работающих с аудиоданными. Это жизнь компьютерного слушания за пределами лаборатории: интернет-боты, скачивающие все доступные им аудио-данные, компьютерные алгоритмы, выделяющие отдельные черты; другие алгоритмы, комбинирующие их для создания паттернов. Дизайн аудио-представления свойств может быть воспроизведен на микроуровне, но мы склонны к экстремумам. Сам принцип компьютерного слушания предвосхищает эту ситуацию: нам нужны компьютеры, которые слушали бы за нас, потому что желание прослушать *всю* музыку невыполнимо.

На этом уровне тардовская связь между статистикой и ощущением появляется снова: статистические методы, появившиеся в социальных науках, находят новое применение в классификации и реорганизации огромного объема информации. Статистические алгоритмы учитывают и поведение пользователей, соединяя социальность слушателей с социальностью (в тардовском смысле) больших наборов данных.

На любом уровне, от двадцатимиллисекундного MFCCs до классификатора, который организует 20 миллионов песен в соответствии с историями прослушивания, алгоритмы работают на данных других алгоритмов. Пока неясно, в какой момент мы можем сказать, что эти алгоритмы перестают представлять распознавание (hearing) и начинают социальность. Если следовать Тарду, такого момента вообще не существует. Мир погружен в колебания и резонанс на всех уровнях, от сложных поведенческих структур big data до разнородных массивов музыкальных колебательных систем. Тардеанский подход предполагает, что мы ищем резонанс между пониманием звука и способами понимания людей. При этом важно уделять внимание работе преобразования между социальным, звуковым и числовым.

Библиография / References:

Helmholtz H. von (1995) On the Physiological Causes of Harmony in Music. *Science and Culture: Popular and Philosophical Essays*, ed. David Cahan. Chicago: University of Chicago Press: 46-75.

Helmreich S. (2008) An anthropologist underwater: Immersive soundscapes, submarine cyborgs, and transductive ethnography. *American Ethnologist* 34(4): 621-641.

Latour B. (2010) Tarde's Idea of Quantification. *The Social After Gabriel Tarde: Debates and Assessments*, ed. Matei Candea. New York: Routledge: 145-162.

Sterne J. (2003) *The Audible Past*. Durham, NC: Duke University Press.

Sterne J. (2012) *MP3: The Meaning of a Format*. Durham, NC: Duke University Press.

Tarde G. (1903) *The Laws of Imitation*, New York: Henry Holt and Company.

Tarde G. (1969) Basic Principles. *On Communication and Social Influence*, ed. Terry N. Clark. Chicago: University of Chicago Press: 143-148.

Рекомендация для цитирования / For citations:

Сивер Н. (2018) Тардовская статистика, машинный слух и Big Data. Все, что воспринимается, — лишь число. *Социология власти*, 30 (3): 193-200.

Seaver N. (2018) Every Sensation Is Only a Number: Tardean Statistics, Computer Audition, and Big Data. *Sociology of Power*, 30 (3): 193-200.

Поступил в редакцию: 24.09.2018; принят в печать: 30.09.2018