Компьютерный анализ звуковысотной системы голоса

Харуто А.В.,
кандидат технических наук

Интонационная составляющая речи, физическим носителем которой является мгновенная частота основного тона (ЧОТ), давно привлекает внимание исследователей как существенная психофизиологическая характеристика (см., например, [Lieberman, 1961; Женило, 1988, 1995]) и как филологический феномен [Кантер, 1988]. Анализ мелодического рисунка вокальной речи позволяет выделять «типовые» фрагменты исполнения — тоны, глиссандо, вибрато — и исследовать их характеристики [Харуто, 1998, 2005; Харуто, Смирнов, 1999; Смирнов, Харуто, 2000].

Анализ звукоряда на основе фонограммы предполагает проведение звуковысотной расшифровки, т. е. построения мелограммы (аналог контура ЧОТ; для удобства музыковедов мелограмма отображает ЧОТ в координатах высоты звука, а не частоты), с последующим её исследованием. Под звукорядом понимается набор звуков определённой высоты, на основе которых построена соответствующая музыкальная система. При исследовании предполагается, что наличие звукоряда проявляется в «более длительном» пребывании ЧОТ звука на определённых этим звукорядом уровнях, в то время как другие значения ЧОТ появляются в фонограмме только кратковременно — при переходе между частотами, относящимися к звукоряду. Выявление частот звукоряда возможно на основе одномерной плотности распределения ЧОТ: в соответствии с принципом максимального правдоподобия положения вершин локальных максимумов, распределения должны совпадать с частотами, образующими звукоряд.

Один пример мелограммы такого рода показан на рис. 1. В программе анализа музыкального звука SPAX, разработанной автором[1], предусмотрен режим отображения «сетки» звуковысотных ступеней при произвольном выборе их числа в октаве, а также возможность подстройки всей «сетки» по высоте; в данном случае наилучшим получилось совпадение высот, на которых «останавливается» голос, примерно с 19-ступенным равномерно темперированным звукорядом (т. е. содержащим 19 эквидистантных ступеней высоты в октаве).

Визуальный анализ характера контура ЧОТ достаточен для предварительных оценок и позволяет понять временную и звуковысотную структуру исследуемого процесса, однако потребность в более объективных данных заставляет разрабатывать алгоритмы анализа, дающие числовую оценку характеристик процесса.

[1] Программа SPAX. Свидетельство ФГУ «Роспатент» о регистрации № 2005612875 от 7 ноября 2005 г.

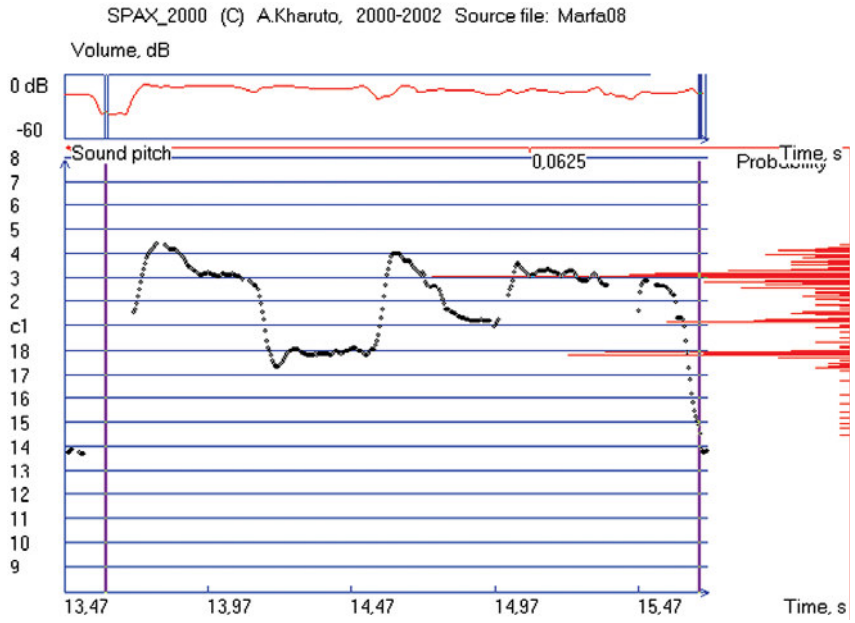

Рис. 1. Мелограмма фольклорного исполнителя, использующего равномерно-темперированный звукоряд примерно с 19-ю ступенями в октаве; диаграмма справа показывает результат анализа распределения «времени пребывания» звука на разных высотах

В примере на рис. 1 показана плотность распределения высоты звука для фрагмента фонограммы, выделенного вертикальными маркерами (график справа; ось вероятности направлена справа налево). Распределение имеет явно выраженные максимумы на тех «привычных» высотах голоса, где наблюдаются длительные горизонтальные участки в мелограмме.

Исследования распределения величины ЧОТ в речевых образцах показали, что оно часто оказывается полимодальным; в работе [Женило, 1988] отмечалось, что у некоторых дикторов моды распределения образуют систему, совпадающую по структуре с равномерно-темперированным музыкальным строем. Как показал ряд исследований автора доклада (см., напр., [Харуто, Смирнов, 1999; Смирнов, Харуто, 2000]), народные фольклорные певцы[2] обычно используют свой индивидуальный звукоряд с интервалом между звуками, меньшим, чем 1/12 октавы (практически от 1/17 до 1/30 и менее).

Следует отметить, что более полные данные могли бы быть получены путём исследования многомерных распределений, учитывающих статистические связи между высотами соседних звуков. Такие зависимости прослеживаются, например, в музыкальном исполнении на инструментах с нефиксированной настройкой: высота изменяется исполнителем по сравнению с «предписанной» нотами величиной для большего благозвучия (т. е. для исправления погрешностей 12-полутонового равномерно-темперированного строя). Измерения, подтверждающие это, были проведены разными исследователями и описаны, например, в работах [Рабинович, 1932; Сахалтуева, 1960; Рагс, 1970].

Очевидно, что наличие звуковысотного вибрато и случайные или преднамеренные неточности выдерживания высоты «размывают» линию, соответствующую положению зву-

[2] Мы сознательно отличаем их от *профессиональных* исполнителей фольклора, которые часто имеют современное музыкальное образование и поют в 12-полутоновом равномерно-темперированном строе.

ковысотной ступени. При анализе фольклорных вокальных фонограмм, где вибрато отсутствует или появляется весьма редко (что можно проконтролировать путём просмотра всей звуковысотной расшифровки типа представленной на рис. 1), непосредственный анализ статистического распределения высоты звука будет давать необходимый результат; в случае более частого использования вибрато может быть произведено предусмотренное в программе SPAX интерактивное измерение каждого тона (т. е. звука с постоянной высотой) и тона, сопровождаемого вибрато. При этом фиксируется среднее значение тона на заданном интервале времени, а также параметры вибрато [Харуто, 2005], и дальнейшее статистическое исследование проводится по этим данным. Ниже мы ограничимся исследованием фонограмм, в которых отсутствует преднамеренное вибрато; будут рассмотрены методы и результаты анализа звуковысотной системы в фольклорном пении, близком к речитативу; для проверки и отладки алгоритмов использован образец фонограммы с музыкально-инструментальным исполнением.

В разработанной автором программе SPAX для определения ЧОТ применяется метод кепстра. Экспериментальная оценка точности определения ЧОТ по синтетическому сигналу показала отклонение от заданной частоты в пределах примерно в 4–5 центов (напомним, что октава составляет 1200 центов, а стандартный полутон равен 100 центам). Использование программы для звуковысотной расшифровки нескольких десятков образцов вокального фольклора не выявило никаких разночтений по сравнению со слуховым анализом фонограмм, проводившимся экспертами-фольклористами.

При исследовании распределения ЧОТ гистограмма строилась из «окон» размером $\Delta h = 5$ центов и (иногда) более. Известно, что размер окна гистограммы влияет на точность её оценивания. Чем меньше размер окна, тем выше точность оценки позиционирования элементов распределения (напр., требуемых в нашем случае локальных максимумов), т. е. меньше систематическая погрешность оценки с помощью гистограммы, использующей замену истинного распределения $\Psi(h)$ системой из N_H прямоугольных окон шириной Δh. Однако уменьшение размера окна приводит к меньшему числу зарегистрированных в нём значений процесса, т. е. меньшей вероятности p_i пребывания процесса в пределах этого i-го окна, что, в свою очередь, приводит к увеличению относительной среднеквадратичной погрешности оценивания значения $\psi(h_i)$ при данной высоте звука h_i, определяемой как (см., например, [Мирский, 1972, с. 313]):

$$\varepsilon^2 = \frac{1}{N} \times \frac{1 - p_i}{p_i},$$

где N — число некоррелированных выборок процесса.

Для выявления локальных максимумов распределения, соответствующих «привычным» частотам исследуемого голоса, могут быть использованы разные подходы. В частности, можно пытаться непосредственно зафиксировать локальные максимумы плотности распределения высоты звука (напомним, что высота пропорциональна логарифму ЧОТ). Для определения точек максимумов следует отыскивать в гистограмме $\{p_i, i = \overline{1, N_H}\}$ точки p_i, возвышающиеся над соседними, т. е. отвечающие условиям

$$p_{j-1} < p_j \quad \text{и} \quad p_j > p_{j+1}.$$

Поскольку положение максимума никак не привязано к границам окон гистограммы, для более точного определения его истинного положения целесообразно использовать аппроксимацию формы кривой $\Psi(h)$ в районе максимума. Например, в одном из исследованных нами алгоритмов через каждые три точки в районе максимума проводилась квадратичная парабола и далее аналитически определялось положение максимума. Очевидное ограничение на возможное расстояние между соседними обнаруживаемыми ступенями состоит в том, что этот интервал не может быть меньше $2 \times \Delta h$, что при $\Delta h = 5$ центам даёт величину минимального фиксируемого «шага» звуковысотной системы в 10 центов.

Такой алгоритм поиска обнаруживает, однако, все «выступающие» точки гистограммы, которых при анализе реального исполнения оказывается очень много и которые, по всей видимости, не являются ступенями звукоряда. Большая часть интервалов между «ступенями» при этом только ненамного превышает указанный нижний предел. Для примера рассмотрим анализ одной музыкальной фонограммы — это упражнение, «коряво» исполненное начинающим скрипачом (он проигрывал гаммы вверх и вниз). «Идеал», к которому стремился исполнявший, — ряд равноотстоящих по высоте «ступенек» на высотах, соответствующим нотам 12-полутонового равномерно-темперированного строя. Однако поскольку скрипка — инструмент с нефиксированной настройкой, здесь возможны (и реально присутствуют) погрешности интонирования. На рис. 2 показан фрагмент мелограммы и гистограмма распределения высот, оценённая по всей фонограмме. Здесь хорошо видна система «пиков» распределения, которые практически не перекрываются, но разнонаправлено сдвинуты по сравнению со стандартными высотами нот 12-полутонового звукоряда. «Пики» распределения имеют также разную ширину и иногда раздвоены, что объясняется нестабильностью высоты при исполнении — «дрожанием» в процессе исполнения одного звука (увеличенная ширина) и неточностью средней высоты при повторном проигрывании той же ступени (раздвоение).

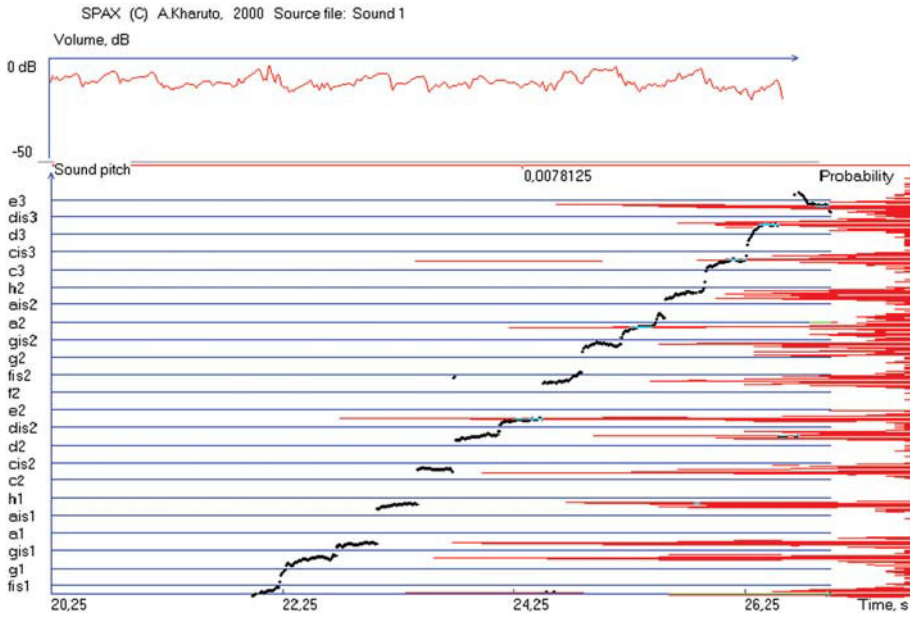

Рис. 2. Фрагмент звуковысотной расшифровки ученического исполнения на скрипке и оценка распределения высот

Рис. 3. Результат оценки звукоряда (см. пример на рис. 2) путём поиска всех локальных максимумов

Определение ступеней звукоряда путём фиксации всех локальных максимумов даёт результат, показанный на рис. 3 (за «нуль» высоты принята нота *до* первой октавы).

Здесь видны как «повторяющиеся» ступени, разделённые очень малыми интервалами (соответствующие, видимо, раздвоенным пикам распределения), так и переменные по величине «скачки» между ступенями. Присутствуют, соответственно, как очень мелкие шаги (соответствующие «сдвоенным» пикам), так и близкие к ожидаемым для данного случая, т. е. примерно кратные 100 центам.

Если об исследуемом звукоряде нет априорных сведений, то по подобным данным определить его структуру было бы затруднительно. Выделение «основных» ступеней по признаку наибольшей вероятности некорректно, поскольку суммарное время пребывания высоты звука в том или другом окне гистограммы, т. е. оценка вероятности «использования» каждой из искомых ступеней, существенно зависит и от исполняемой мелодии (а в случае речевого общения — от требуемой интонации высказывания), и от «качества» её воспроизведения, что иллюстрируется рис. 2. Таким образом, вполне возможно, что коротко прозвучавший тон окажется одним из основных, образующих звуковысотную систему, так что его нельзя не учитывать. Кроме того, в исполнении (фрагменте речи) могут отсутствовать некоторые ступени звукоряда, поскольку они «не нужны» в данном случае (но понадобятся в другом).

Используя предположение о том, что искомый звукоряд является равномерно-темперированным, т. е. что его ступени разнесены на равные интервалы по шкале высоты, можно предложить другой способ анализа, основанный на оценке всей гистограммы в целом и поиске в ней *периодически повторяющихся «пиков»*. Для такой оценки автором был (как и для определения ЧОТ) использован метод кепстра: гистограмма логарифмировалась (что

Рис. 4. Спектр, вычисленный для распределения высот звука (см. пример на рис. 2)

«уравнивает» в некоторой степени вклад в оценку часто и редко используемых ступеней), затем вычислялся её спектр. Для приведённого выше примера — ученического исполнения на скрипке — получается спектр распределения, показанный на рис. 4.

Наиболее мощные максимумы обнаруживаются в точках №№ 5, 7, 11 и 15; интенсивности соответствующих компонент спектра отображают «выраженность» данной периодической составляющей (т. е. совокупности «пиков», размещённых через соответствующий шаг по высоте). График интенсивностей для перечисленных наиболее выраженных периодических составляющих в зависимости от предполагаемых шагов между ступенями звукоряда показан на рис. 5. Здесь видно, что наиболее выраженной (наиболее вероятной) структурой в данном исполнении является звуковысотная система с шагом 100 центов (которая и «предписана» стандартными высотами нот 12-полутонового звукоряда). Ошибки исполнения порождают побочные пики, но их интенсивность намного меньше.

Рис. 5. Зависимость «выраженности» периодической компоненты звукоряда от шага между ступенями (см. пример на рис. 2)

Рис. 6. Спектр, вычисленный для распределения высот звука в фонограмме русского фольклорного пения (см. рис. 1)

Анализ первого из приведённых выше примеров (русское фольклорное пение в звуковысотной системе, содержащей, по предварительной оценке, примерно 19 ступеней в октаве, — см. рис. 1), даёт спектр плотности распределения, приведённый на рис. 6. На рис. 7 показаны значения интенсивности периодической компоненты распределения высоты для основных «пиков» данного спектра. Здесь число значительных и сопоставимых по величине максимумов значительно больше, однако для самого мощного из них (точка № 18) получается величина соответствующего шага по высоте, равная 58,8 центам, что соответствует 1200:58,8=20,4 ступеням в октаве. Другие пики спектра получаются из-за большого количества промежуточных по высоте звуков — в частности, из-за украшения мелодии движением вверх-вниз относительно средней высоты исполняемого звука.

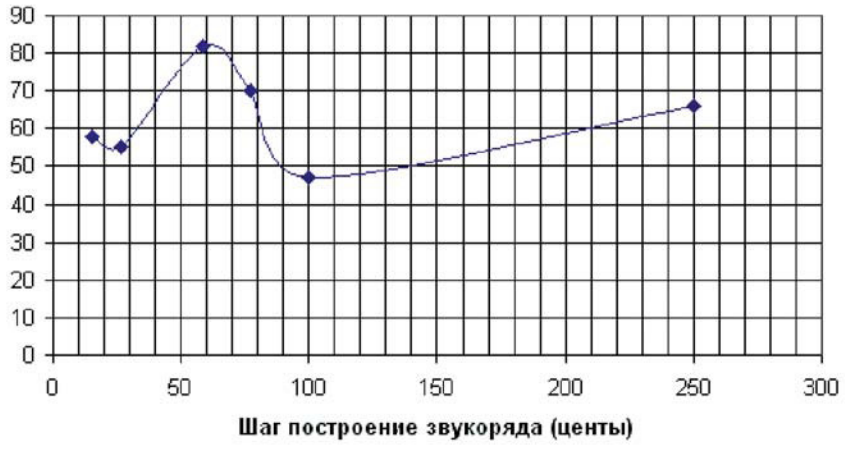

Рис. 7. Зависимость «выраженности» периодической компоненты звукоряда от шага между ступенями для образца русского фольклорного пения (см. рис. 1)

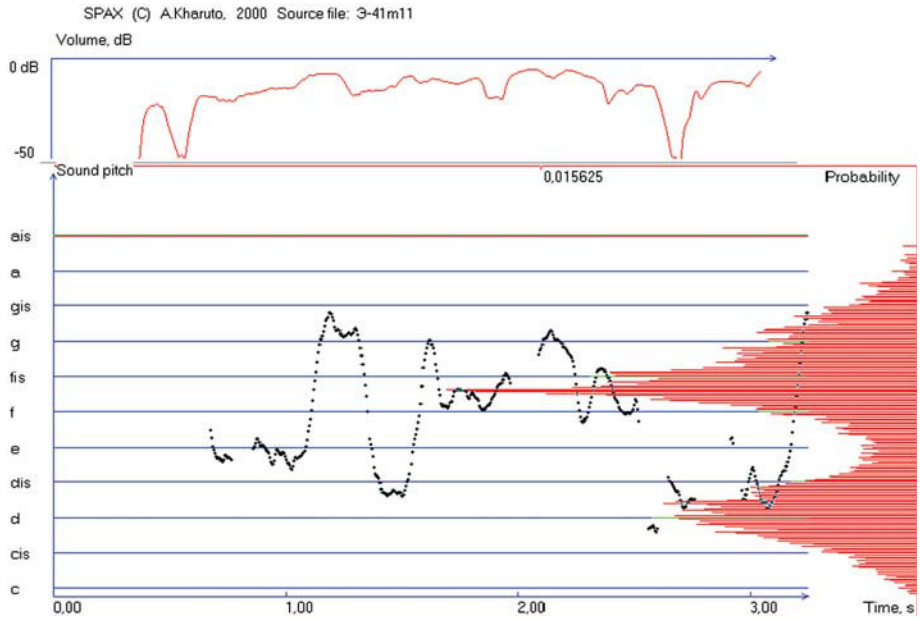

Рис. 8. Фрагмент звуковысотного рисунка и распределение высоты в эвенкийской песне

Ещё один пример анализа — эвенкийская песня (мужской голос, исполнение близко к речитативному). Звуковысотная расшифровка фрагмента и гистограмма распределения высоты показана на рис. 8. Спектр для этого распределения показан на рис. 9. Как и в предыдущем случае, спектр содержит много максимумов, что отображает сложную звуковысотную структуру[3]. На рис. 10 показаны интенсивности для наиболее выраженных периодических составляющих этого спектра.

При анализе спектра распределения высоты для этого образца выявляется основной по выраженности шаг звуковысотной системы, равный 20 центам (точка № 51 в спектре).

Таким образом, компьютерный анализ звуковысотной системы фонограммы на основе распределения высоты звука с последующим исследованием периодичности структуры этого распределения позволяет определить интервал, образующий равномерно-темперированный звукоряд исполнителя. По-видимому, можно также на основе характера спектра распределения (количества дополнительных «пиков», их интенсивности и пр.) интегрально оценивать точность следования звукоряду в исследуемом исполнении.

Отметим, что указанный тип звукоряда не является единственно возможным: так, в тувинском горловом пении (и сходных с ним монгольском, тибетском и др.), где во время вокализов слышны по меньшей мере одновременно два голоса на разных высотах, мелограмма «верхних» голосов показывает использование *натурального* звукоряда, где звуковысотные ступени разнесены на равные *по частоте* (а не по высоте) интерва-

[3] Здесь (как и на других графиках спектров) для удобства масштабирования «вырезана» постоянная составляющая, вследствие чего образовался искусственный максимум в точке № 2, — в расчётах он не учитывается.

Рис. 9. Спектр, вычисленный для распределения высот звука в фонограмме эвенкийского фольклорного пения (см. рис. 8)

лы. Это связано с используемым механизмом звукоизвлечения: все слышимые «голоса» образуются из обертонов нижнего основного звука, т. н. *бурдона*, высота которого во время исполнения практически неизменна (см, например, [Харуто, Карелина, 2008; Харуто, 2008].

Сопоставляя результаты нашего анализа с традициями европейского этномузыковедения, использующего 12-полутоновую нотацию (иногда — с дополнительными знаками микроальтерации, позволяющими фиксировать, например, четвертитоновые интервалы), можно заключить, что зарегистрирован-

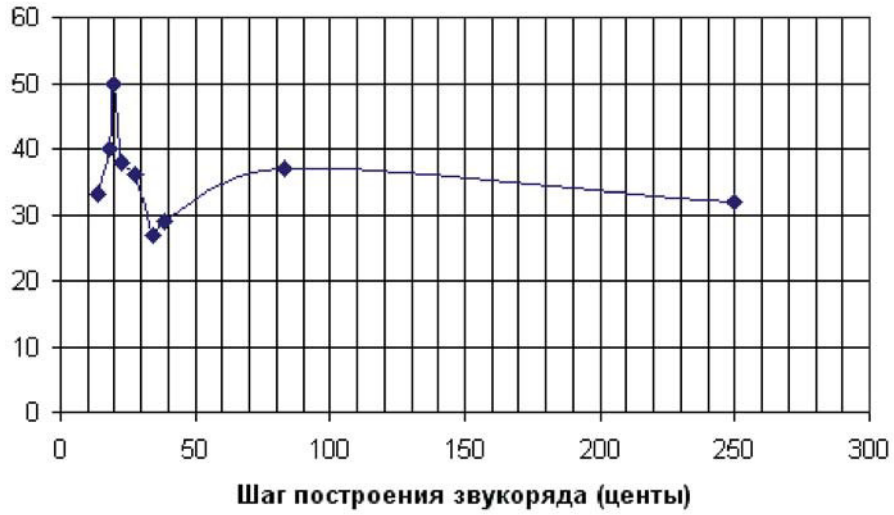

Рис. 10. Зависимость «выраженности» периодической компоненты звукоряда от шага между ступенями для образца эвенкийского фольклорного пения (см. рис. 8)

ные в фольклорном пении интервалы между ступенями должны измеряться гораздо точнее и не могут быть отображены указанными средствами нотации.

Следует отметить, что представленные результаты носят предварительный характер и требуют дальнейшей проверки и сопоставления с данными, полученными экспертами-музыковедами с помощью «традиционных» слуховых методов.

Литература

1. *Lieberman Ph.* (1961) Perturbations in Vocal Pitch // The Journal of the Acoustic Soc. of America. 1961. v.33, N 5. — p. 597–603.
2. *Женило В. Р.* Анализ параметров частоты основного тона голоса человека для автоматической идентификации личности // Академия наук СССР, Вычислительный центр. Сообщения по программному обеспечению ЭВМ. М., 1988.
3. *Женило В. Р.* Компьютерная фоноскопия // М: Академия МВД России, 1995.
4. *Кантер Л. А.* Системный анализ речевой интонации: Учебн. пособие. М.: Высшая школа, 1988.
5. *Харуто А. В.* Компьютерный анализ звука в музыковедческом исследовании. Труды международного научного симпозиума «Информационный подход в эмпирической эстетике». Таганрог: Изд. ТРТУ, 1998.
6. *Харуто А. В., Смирнов Д. В.* Использование компьютерного анализа в исследовании звуковысотного строения народной музыки //Материалы международных конференций памяти А. В. Рудневой. М.: Московская гос. консерватория, 1999.
7. *Смирнов Д. В., Харуто А. В.* Нелинейный звукоряд в музыкальном фольклоре: общая закономерность и индивидуальность // Языки науки — языки искусства // Общ. ред. З.Е. Журавлевой, В. А. Копцик, Г. Ю. Резниченко. М.: МГУ, 2000.
8. *Харуто А. В.* Статистическое исследование характеристик вибрато // Сборник трудов XIV международной научной конференции «Информатизация и информационная безопасность правоохранительных органов». М.: Академия управления МВД России, 2005.
9. *Рабинович А. В.* Осциллографический метод анализа мелодии // Проблемы музыкознания. Теоретическая библиотека. М.: Музгиз, 1932.
10. *Сахалтуева О. Е.* О некоторых закономерностях интонирования в связи с формой, динамикой и ладом // Труды кафедры теории музыки Московской гос. консерватории им. П. И. Чайковского. Вып. 1. М.: Музгиз, 1960.
11. *Рагс Ю. Н.* О художественной норме чистой интонации при исполнении мелодии: Дисс. канд. искусствоведения. М., 1970.
12. *Мирский Г. Я.* (1972) Аппаратурное определение характеристик случайных процессов. Изд. 2-е перераб. и доп. М.: Энергия, 1972.
13. *Харуто А. В., Карелина Е. К.* К вопросу о музыкально-акустических свойствах тувинского горлового пения. «Музыкальная академия», 2008. № 4, С. 108–113.
14. *Харуто А. В.* Тувинское горловое пение: акустический анализ и модель звукообразования // Сб. трудов XX сессии Российского акустического общества, секция «Акустика речи» М., РАО, 2008. С. 106–110.

Харуто А.В.
Кандидат технических наук,
доцент Московской государственной консерватории.