

КРАТКИЙ ОБЗОР МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ ЛОКАЛИЗАЦИИ ЗВУКА В ВИРТУАЛЬНОЙ СРЕДЕ

А.Ю. Разваляева

Институт психологии РАН

Россия, 129366, Москва, Ярославская ул., 13

E-mail: annraz@rambler.ru

Ключевые слова: слуховое восприятие, моделирование, локализация звука, кажущийся источник звука, передаточная функция головы (HRTF).

Аннотация: Локализация звука в пространстве – одна из фундаментальных характеристик слухового восприятия. Актуальной задачей выступает поиск таких способов передачи пространственного слуха в виртуальных средах, которые будут наилучшим образом восприниматься слушателем. Рассматриваются математические модели локализации звука в пространстве, входящие в библиотеку The Auditory Modeling Toolbox, а также метрики перцептивных ошибок, позволяющие проводить дополнительную оценку и сравнение моделей, и модели, в которых локализация звука рассматривается на уровне подпроцесса. Приводятся примеры научных и практических задач, которые могут решаться с помощью моделей локализации звука. В качестве перспективы данного направления выделяется построение более экологически валидных моделей, позволяющих оценивать восприятие сложных звуков.

1. Введение

Локализация источника звука в пространстве – неотъемлемая функция слухового восприятия, участвующая в процессах выделения речи в шуме (эффект вечеринки с коктейлями), распознавания сложных акустических событий, формирования представлений об окружающих предметах и пространстве (например, о размерах комнаты). Механизмы локализации звука достаточно подробно изучены и включают интерауральные временные различия (*Interaural Time Delay*), интерауральные различия по интенсивности (*Interaural Intensity Difference*), спектральные изменения звука, зависящие от индивидуально-анатомических особенностей строения ушной раковины, головы и торса, изменение характеристик звука во время движения человека или источника звука, контекст прослушивания (прошлый опыт, влияния со стороны зрительного восприятия) [1]. Однако остаются неизученными, с одной стороны, вопросы локализации сложных звуков в экологически валидных ситуациях. С другой стороны, перспективной областью исследований выступает воссоздание пространственных характеристик звука в виртуальной среде, что может применяться для более легкого опредмечивания акустического события, повышения воспринимаемого качества звука, использования акустической информации в задачах управления системой [2].

1.1. Техники обработки звука в виртуальной среде

Звук в виртуальной среде передается слушателю через наушники и без специальной обработки воспринимается как локализованный не в пространстве вокруг слушателя, а звучащий внутри его головы (т.е., не происходит «экстернализация»

звука) [3]. Для передачи пространственных свойств может использоваться передаточная функция головы (*Head-Related Transfer Function*; HRTF), получаемая с помощью записи звука из множества положений в пространстве на микрофоны, которые располагаются в ушной раковине человека. HRTF позволяет передать индивидуальные анатомические особенности, связанные с восприятием положения звука. Выделяются HRTF двух типов – индивидуальные и неиндивидуальные; последние, как правило, записываются на манекене. Обработка звука с помощью неиндивидуальных HRTF приводит к повышению ошибок локализации звука, особенно тех, которые представляют собой «переворачивание» оси – ошибки «фронт-тыл» или «верх-низ» [1].

Кроме HRTF для создания кажущегося источника звука, локализованного вне головы, рекомендуется использовать также многоканальные записи, сводящиеся на два канала (амбисоник), и импульсные отклики помещения для передачи информации о реверберации [3]. Амбисоник также включает несколько разновидностей (порядков), отличающихся количеством «сворачиваемых» каналов изначальной многоканальной записи и качеством получаемых звуков.

2. Модели локализации звука

2.1. Математическое моделирование локализации звука на примере The Auditory Modeling Toolbox

Рассмотрим основные направления моделирования локализации звука в виртуальной среде на примере библиотеки для моделирования слухового восприятия для MatLab/Octave (*The Auditory Modeling Toolbox*). На настоящий момент библиотека содержит 19 моделей локализации звука в пространстве, большая часть которых проверена и воспроизводится на различных данных [4]. Моделируются психофизиологический отклик слуховой системы на звук или воспринимаемое качество звука (направление, расстояние, ширина кажущегося источника звука) [5]. Все модели (и функции, с помощью которых они вызываются) названы по фамилии первого автора модели и году ее публикации. Представленные в документации к библиотеке [6] модели включают:

- Модели латерализации – восприятия звука справа или слева (dietz2011, laback2023), в том числе у пользователей кохлеарных имплантов (kelvasa2015). Отдельно можно выделить модель латерализации, основанную на алгоритмах машинного обучения (обучения с учителем – may2013). Все эти модели подсчитывают интерауральные временные различия и интерауральные различия по интенсивности, а также содержат несколько последовательных этапов, на которых применяются фильтры, имитирующие работу слуховой системы.
- Модели локализация в разных плоскостях – сагиттальной (baumgartner2013, baumgartner2014, baumgartner2016) и медианной (langendijk2002). Данные модели принимают бинауральные импульсные отклики, специфичные для слушателя, и возвращают вероятностное предсказание угла ответа (а функция baumgartner2013 может также использовать фильтры для моделирования периферической обработки звукового сигнала). При этом информация о плоскости выделяется из импульсных откликов или предоставляется отдельно.
- Модель восприятия расстояния (georganti2013) принимает на вход характеристики бинаурального звука, которые подаются на классификатор в модель Гауссовой смеси.

- Байесовские модели локализации трехмерного звука (barumerli2023, mclachlan2021, geijniers2014) основываются на HRTF слушателя, импульсных откликах или интерауральных временных различиях и предсказывают воспринимаемое направление на кажущийся источник звука и расхождения с целевым направлением.
- Модели экстернализация звука на основе интерауральных различий по интенсивности (hassager2016) и импульсных откликов помещения для имитации пространства с реверберацией (li2020).
- Модель подсчета различий между воспринимаемым и желаемым направлением на источник звука в стереофонических системах (wierstorf2013).
- Нейронная модель локализации в условиях, когда слушатель носит головной убор или устройство для защиты слуха (llado2022).

2.2. Математические модели слухового восприятия, связанные с локализацией

Можно выделить класс моделей, включающих в себя репрезентации определенных механизмов локализации звука, но направленных на другие феномены слухового восприятия.

Модели бинаурального слуха могут включать функции для репрезентации психофизиологии слуховой системы (напр., улитки, медиальной и латеральной верхней оливы – takanen2013), интерауральных временных различий и интерауральных различий по интенсивности (lindemann1986), интерауральной когерентности (согласованность сигнала, приходящего на левое и правое ухо – eurich2022) и импульсных откликов помещения (bischof2023). Некоторые модели бинауральной детекции речи в шуме также опираются на принцип разделенности речи и шума в пространстве и включают импульсные отклики головы (jelfs2011) и импульсные отклики помещения (leclere2015) [6].

Отдельно можно выделить модели, основывающиеся на теории анализа слуховых сцен и модели слухового внимания. Такие модели включают процессы двух типов: bottom-up (процессы восприятия, психофизиологические процессы) и top-down (когнитивные процессы). Хотя эти модели часто игнорируют пространственные характеристики звука, феномены, которые они призваны объяснить (различение речи и игнорирование шума, произвольное внимание к необычным стимулам, сегрегация слуховых потоков), могут быть связаны с локализацией источника звука в пространстве [7].

2.3. Перцептивные метрики

В исследованиях, моделирующих локализацию звука, также применяются дополнительные функции для обозначения перцептивных ошибок. Чаще всего моделируются полярные ошибки – отклонения от целевого положения в пределах одной полусферы (метрика точности ответов), и ошибки квадранта – ошибки, превышающие 90°, например, ошибки фронт-тыл. Более частные метрики – глобальная полярная ошибка – метрика, «игнорирующая» ошибки фронт-тыл, и частота ошибок фронт-тыл – частота ответов, в которых звук был локализован спереди, когда его источник находился сзади и наоборот, без учета углов подъема. Применение более частных метрик позволяет сравнивать модели и способы репрезентации пространственного звука в виртуальных пространствах с учетом данных о высокой частоте ошибок фронт-тыл в некоторых частных случаях (например, при применении неиндивидуализированных HRTF) [8].

2.4. Применение моделей для решения научных и практических задач

Приведенные выше модели чаще всего используются для решения научных задач, например, для перепроверки и обобщения результатов экспериментальных исследований и выдвижения новых гипотез о работе слуховой системы [5]. Так, модели локализации звука в сагиттальной плоскости позволили провести сравнение различных алгоритмов записи HRTF, использовавшихся в психоакустических лабораториях по всему миру, без привлечения реальных испытуемых, и показали существенные различия между разными базами HRTF по количеству ошибок локализации [9]. Однако при таком использовании моделей необходимо дополнительно решать проблему валидности и надежности получаемых результатов, дисперсии перцептивной неуверенности, измеряемой моделью, и соответствия результатов тем, которые подтверждаются экспериментальными исследованиями на людях [8, 9].

В практических целях модели локализации звука должны быть адаптированы для нужд конкретной области применения: так, в клинических исследованиях в модель может вводиться поражение определенных участков слуховой системы [5]. В области исследований городской среды предлагается перейти от моделей шума к динамическим статистическим моделям пространственного звука, которые могут использоваться для сохранения или улучшения звуковой среды при создании новых городских сооружений [10]. В играх модели пространственного слуха используются для наделения неигровых персонажей, управляющихся искусственным интеллектом, большей реалистичностью и привлекательностью для пользователя [11].

3. Заключение

Большая часть моделей слуховой локализации на настоящем этапе строится на основе представлений о психофизиологии слуховой системы, а также данных психоакустических экспериментов. При этом встает задача сравнения качества моделей (валидности, воспроизводимости результатов) и их пригодности для решения различных задач. Математические модели чаще всего основываются на данных лабораторных исследований и настраиваются на воспроизведение ответов «идеального испытуемого», в силу чего затруднительно оценить, насколько их результаты применимы к более экологически валидным ситуациям. Перспективным видится построение моделей на основе экспериментов, включающих качественную оценку воспринимаемого качества события [2].

Исследование выполнено по Государственному заданию № 0138-2024-0012.

Список литературы

1. Разваляева А.Ю., Носуленко В.Н. Пространственная локализация цифрового звука в научном эксперименте и практике // *Экспериментальная психология*. 2023. Т. 16, № 2. С. 20-35. DOI: <https://doi.org/10.17759/expsy.2023160202>.
2. Носуленко В.Н., Харитонов А.Н. Жизнь среди звуков. Психологические реконструкции. М.: Институт психологии РАН, 2018. 422 с.
3. Алдошина И.А., Игнатов П.В., Иванов Ю.М. Бинауральный синтез в искусстве записи и воспроизведения звука // *Современные проблемы науки и образования*. 2015. № 1. <https://science-education.ru/ru/article/view?id=17467> (дата обращения: 02.02.2023).
4. <https://www.amtoolbox.org/models.php> (дата обращения: 19.01.2024).
5. Søndergaard P.L., Majdak P. The Auditory Modeling Toolbox // Blauert J. (ed.) *The Technology of Binaural Listening*. Berlin, Heidelberg: Springer, 2013. P. 33-56. DOI 10.1007/978-3-642-37762-4.
6. <https://www.amtoolbox.org/amt-1.5.0/doc/models/index.php> (дата обращения: 19.01.2024).

7. Kaya E. M., Elhilali M. Modelling auditory attention // *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2017. Vol. 372, No. 1714. P. 20160101.
8. Barumerli R., Geronazzo M., Avanzini F. Localization in elevation with non-individual head-related transfer functions: Comparing predictions of two auditory models // *2018 26th European Signal Processing Conference (EUSIPCO)*. IEEE, 2018. P. 2539-2543.
9. Barumerli R., Geronazzo M., Avanzini F. Round Robin Comparison of Inter-Laboratory HRTF Measurements—Assessment with an auditory model for elevation // *2018 IEEE 4th VR workshop on sonic interactions for virtual environments (SIVE)*. IEEE, 2018.
10. Hong J.Y., Jeon J.Y. Exploring spatial relationships among soundscape variables in urban areas: A spatial statistical modelling approach // *Landscape and Urban Planning*. 2017. Vol. 157. P. 352-364.
11. Cowan B., Kapralos B., Collins K. Spatial Audio Modelling to Provide Artificially Intelligent Characters with Realistic Sound Perception // *Millennium Biltmore Hotel, Los Angeles, CA*. 2016. P. 3.