МИКРОФОННЫЕ РЕШЕТКИ, УЛУЧШЕННЫЙ ПРИЕМ РЕЧИ И ИХ ДОСТИЖЕНИЯ В СФЕРЕ ОБРАБОТКИ СИГНАЛОВ

Рагачурина Н.А.
Ташкентского университета информационных технологий имени Мухаммада ал-Хоразмий
n.ragachurina@tuit.uz

Аннотация: *в статье представлен всесторонний обзор эволюции технологий микрофонных решеток и многоканального улучшения речи — от исторических истоков до современных коммерческих решений. В работе представлена классификация базовых методов проектирования акустических систем, алгоритмы борьбы с шумом и реверберацией, а также передовые подходы на стыке обработки сигналов и глубокого обучения (включая нейросетевое формирование луча). В работе проанализировано влияние данных технологий на пользовательский опыт, обозначены ключевые технологические вызовы и предложены перспективные направления будущих исследований, способные стимулировать дальнейшие инновации в индустрии.*

Ключевые слова: *микрофонная решетка, передача сигналов, beamformers, обработка речи.*

Annotatsiya: *Ushbu maqolada mikrofon massivi va ko'p kanalli nutqni yaxshilash texnologiyalarining tarixiy kelib chiqishidan tortib, zamonaviy tijorat yechimlarigacha bo'lgan evolyutsiyasi haqida keng qamrovli ma'lumot beradi. Unda asosiy akustik tizimlarni loyihalash usullarining tasnifi, shovqin va aks-sadoni kamaytirish algoritmlari hamda signalni qayta ishlash va chuqur o'rganish (shu jumladan, neyron tarmoq nurlarini shakllantirish) kesishmasidagi ilg'or yondashuvlar keltirilgan. Bunda ushbu texnologiyalarning foydalanuvchi tajribasiga ta'siri tahlil qilinadi, asosiy texnologik muammolar aniqlanadi va sohada keyingi innovatsiyalarni rag'batlantirishi mumkin bo'lgan kelajakdagi tadqiqotlar uchun istiqbolli yo'nalishlar taklif etiladi.*

Kalit so’zlar: *mikrofon massiv, signalnni uzatis, nutqga ishlov berish, beamformers.*

Abstract: *This article provides a comprehensive overview of the evolution of microphone array technologies and multichannel speech enhancement—from their*

historical origins to contemporary commercial solutions. It presents a classification of fundamental acoustic system design methods, algorithms for mitigating noise and reverberation, and advanced approaches that integrate signal processing with deep learning (including neural-network-based beamforming). The paper analyzes the impact of these technologies on user experience, identifies key technical challenges, and proposes promising directions for future research that could drive further innovation within the industry.

Keywords: *microphone array, signal transmission, beamformers, speech processing.*

Введение

В современном мире электроники микрофоны являются неотъемлемой части электрических оборудований. Они применяются в таких системах, как средства принимающие голосовые и речевые сигналы. При разговори или произведения звуков микрофоны улавливают колебание воздуха в виде звуковых волн и преобразует их в электрический сигнал. В дельнейшем этот электрический сигнал может хранится в источнике, может быть обработан и передан. Но на практике не все так просто. При приеме полезного сигнала от источника, кроме нужных нам сигналов принимается и не нужные факторы, такие как посторонние шумы, помехи и нелинейные искажения. А это негативно влияет на принимаемый полезный сигнал – может искажать принимающий сигнал, снижает эффективность работы систем автоматического распознания речи (ASR). Что бы устранить эти проблемы исследователи разрабатывали различные методы улучшения речевого сигнала. Так как первые устройства связи имели лишь один микрофон исследователи были сосредоточены на одноканальных системах связи. Одноканальный метод был эффективным в соотношении сигнал/шум и в приеме качественного сигнала речи, однако проблема было при приеме сигнала с шумом и искажении речевого сигнала [1]. Кроме этих проблем одноканальные микрофоны не различали говорящих и из-за плохого восприятия окружающих звуков не давали глубокое ощущение пространства говорящего.

Вышеуказанные ограничения привели к созданию микрофонных решеток. Микрофонные решетки — это система, где несколько микрофонов объединяются для создания более эффективного восприятия окружения происходящего для получения объёмного звука. Микрофонная решетка обычно состоит из двух основных компонентов: аппаратной части и блока

обработки сигналов. Аппаратная часть включает в себя набор микрофонов и электронные компоненты, такие как предусилители, аналого-цифровые преобразователи (АЦП), блоки памяти и процессоры. Блок обработки сигналов реализует различные алгоритмы, использующие цифровые сигналы с микрофонов для выполнения конкретных задач [2]. Все исследования в основном сосредоточены на разработке различных топологий, то есть на расположении микрофонов внутри системы. Причина такого внимания на топологию микрофонов является попытки достичь эффективного расположения их поскольку конфигурация расположений микрофонов в системе определяет продуктивность системы. На практике выбор топологии решетки может быть ограничен конструкцией устройства, в которое она встраивается; к числу распространенных конфигураций относятся линейные, круговые, концентрические и сферические решетки [2, 3]. Что бы все микрофоны в системе работали исправно надо определенные параметры согласовать. На пример динамический диапазон микрофонов должны совпадать друг с другом. Кроме динамического диапазона имеется не менее значимые параметры такие как чувствительность, амплитудно-частотная характеристика, отношения сигнал/шум и т.д. Конечно же это не единственная цель микрофонов. Основной проблемой микрофонных решёток является качество принимающего звука. Для этого сделаны много исследований, в результате чего улучшился сигнал. В зависимости от характера взаимосвязи между сигналами микрофонов алгоритмы можно разделить на методы, не зависящие от сигнала (например, формирование луча с фиксированными параметрами [3, 4]), и методы, зависящие от сигнала (например, адаптивное формирование луча [5]). С точки зрения стратегий оптимизации моделей сигнала их классифицируют на методы, основанные на моделях (использующие статистические модели и методы оптимизации для улучшения качества речи [4, 6]), и методы, основанные на данных (использующие сами данные для настройки и совершенствования алгоритмов, например, подходы на базе глубокого обучения [7, 8]).

Минимизация микрофонов дает возможность созданию микрофонных решеток, беспроводных акустических сетей микрофонов, которые дают возможность для решений многих задач, где имеется акустические сигналы и применяется в различных областях [9]: слуховые аппараты, сигнализационные системы, конференции и различные современные системы. Помимо этих традиционных сфер использования, микрофонные решетки все чаще применяются в таких областях, как биомониторинг, системы охранного

наблюдения и системы слухового восприятия роботов [10]. За последние десятилетия стремительное развитие технологий микрофонных решеток и методов многоканального улучшения речи привело к росту их значимости и появлению множества новых областей применения [11].

В частности, значительный прогресс в области обучения, достигнутый за последнее десятилетие, способствовал переходу к использованию методов, основанных на данных и технологиях обучения, в системах микрофонных решеток и методах многоканального улучшения речи [12–14].

Микрофонные решетки

Совокупность нескольких микрофонов в одном устройстве называются микрофонными решетками. Для работы таких микрофонов без сбоя должны соблюдаться несколько факторов: они должны располагаться в соответствии определенной конфигурации; иметь совпадающую АЧХ и т.д. Только при придерживании этих условий микрофонная решетка принимает улучшенные сигналы от конкретных источников звука. В радиотехнике до исследований в области микрофонных решеток были популярны разработки в теории антенных решеток. Ранее исследования в этой области во многом опирались на теорию антенных решеток, применяемую в радиолокации и гидролокации. Позднее было ясно, что задачи, решаемые при использовании микрофонных решеток, существенно отличаются от задач радиолокации и гидролокации по ряду важных причин [15]:

• Микрофонные решетки работают с речевыми и аудиосигналами, которые являются широкополосными (диапазон частот примерно от 20 Гц до 20 кГц). Разработка алгоритмов, обеспечивающих стабильную работу в столь широком частотном диапазоне, представляет собой сложную задачу.

• Для работы микрофонных решеток часто требуется оценивать статистические характеристики сигналов. Однако малоподвижность речевых и аудиосигналов значительно усложняет процесс такой оценки.

• Микрофонные решетки часто применяются в условиях реверберации с большим количеством отражений, что может существенно снизить эффективность алгоритмов обработки.

• Из-за физических и вычислительных ограничений, свойственных коммерческим продуктам, количество микрофонов обычно невелико, что затрудняет обеспечение высокой пространственной избирательности.

Вышеуказанные причины, которые подходит в области радиотехники демонстрирующие хорошие результаты в радиолокации и гидролокации, часто показывают недостаточно хорошие результаты в микрофонных

решетках. Поэтому в последнее время исследовании в области создания алгоритмов для микрофонных решеток.

Расположение микрофонных решеток в соответствии их эффективности позволяет применить фильтр выделяющих и улучшающих сигнал от целевого источника. [16, 17]. Из-за аппаратных ограничений в ранних системах формирования луча с использованием микрофонных решеток часто применялся метод «задержка и суммирование» (DS) — простой способ обеспечения направленности [2]. Однако ширина луча в таких системах обратно пропорциональна частоте: основной лепесток диаграммы направленности расширяется на низких частотах и сужается на высоких. Эта особенность снижает эффективность подавления шумов и помех, особенно в низкочастотном диапазоне. Кроме того, частотно-зависимая характеристика DS-формирователя может непреднамеренно действовать как фильтр нижних частот при отклонении угла падения сигнала от направления наведения, а также не обеспечивает равномерного подавления шума во всем частотном спектре. В результате разработка широкополосного формирователя луча, обеспечивающего стабильные характеристики в широком диапазоне частот, представляет собой сложную и трудновыполнимую задачу.

Одним из предложенных решений этой задачи является разделение решетки на подрешетки, каждая из которых рассчитана на работу в определенном частотном диапазоне. В результате формируется микрофонная решетка, состоящая из нескольких гармонически вложенных подрешеток, что позволяет охватить интересующий широкий диапазон частот [2, 18]. Однако такие вложенные подрешетки могут иметь значительные размеры, что не всегда удобно для практического применения. Альтернативным решением является частотно-временная фильтрация на основе фазовых сдвигов [19] с обеспечением постоянной ширины диаграммы направленности [20]; этот метод особенно эффективен для систем, используемых в персональных компьютерах. Еще одним действенным способом обеспечения стабильных характеристик в широкой полосе частот является использование дифференциального диаграммообразователя. Этот подход широко применяется на практике благодаря способности обеспечивать высокий коэффициент направленного действия и формировать частотно-независимые диаграммы направленности при использовании компактных микрофонных решеток [21–23].

Концепция дифференциальных микрофонных решеток (DMA) возникла под влиянием принципа работы направленных ленточных микрофонов,

измеряющих колебательную скорость частиц среды или разность давлений [24, 25]. Существенным недостатком ранних направленных микрофонов была невозможность адаптивной настройки диаграммы направленности в соответствии с меняющимися требованиями: их характеристики направленности, по сути, задавались на этапе производства.

Тем не менее, существовали модели с возможностью выбора различных диаграмм направленности — например, микрофон Western Electric 639B (Birdcage), оснащенный как ленточным, так и динамическим элементами, что позволяло пользователю выбирать один из шести вариантов направленности.

Современная концепция DMA была разработана с применением методов цифровой обработки сигналов: микрофоны, реагирующие на звуковое давление, располагаются в определенном геометрическом порядке, что позволяет измерять разности (дифференциалы) поля звукового давления путем комбинирования выходных сигналов нескольких всенаправленных микрофонов. Также предложен новый метод проектирования дифференциальных диаграммообразователей в области кратковременного преобразования Фурье (STFT) с использованием ограничений типа «нуль» (null-constraints); этот метод обеспечивает значительные преимущества в плане гибкости проектирования и робастности, в частности, позволяя решить проблему усиления белого шума [26, 27]. Были предложены обобщения метода дифференциального диаграммообразования для круговых [15, 28], концентрических круговых [29] и прямоугольных [11] микрофонных решеток, что открывает дополнительные возможности, такие как улучшенное управление направленностью и повышенная робастность системы [30]. Одним из наиболее показательных методов формирования диаграммы направленности (beamforming) в системах с микрофонными решетками является подход, основанный на разложении звукового поля по сферическим гармоникам [31]. Вместо формирования диаграммы направленности путем традиционного интегрирования отфильтрованных сигналов от отдельных микрофонов, в данном методе для проектирования системы формирования луча используются пространственно разложенные собственные лучи (eigenbeams). Сферическая микрофонная решетка Eigenmike® стала важной вехой в развитии технологий микрофонных решеток [32]. Система Eigenmike состоит из двух основных компонентов: формирователя собственных лучей (eigenbeamformer), преобразующего сигналы микрофонов в ортонормированные сферические гармоники, и модального формирователя луча (modal beamformer), который объединяет эти гармоники для создания

требуемой диаграммы направленности. Передовая конструкция и широкие функциональные возможности Eigenmike сделали это устройство ключевым элементом в развитии технологий микрофонных решеток и обеспечили его широкое распространение среди исследователей.

Передовые исследования в области микрофонных решеток

На фоне стремительного развития глубокого обучения — подобно тому, как это происходит в других областях исследований, связанных с обработкой аудио- и акустических сигналов (AASP) [10], — в сфере микрофонных решеток и многоканального улучшения речевого сигнала наблюдается переход к методам, основанным на данных и технологиях глубокого обучения [33]. За последнее десятилетие эта тенденция способствовала значительным достижениям, приведя к разработке множества методов, использующих глубокие нейронные сети (DNN). Такие методы, как использование идеальной бинарной маски (ideal binary mask) и идеальной маски отношений (ideal ratio mask), позволяют формировать маски, указывающие, содержит ли конкретный частотно-временной элемент речевой сигнал или шум; при этом нейронные сети обучаются аппроксимировать эти идеальные маски [34]. С помощью глубоких нейронных сетей (DNN) оцениваются ключевые статистические параметры, необходимые для традиционного формирования луча (beamforming), — например, ковариационная матрица шума или относительные передаточные функции (RTF) полезного сигнала.

DNN повышают точность этих оценок, обучаясь распознавать характеристики шума и сигналов в данных, что, в свою очередь, улучшает работу системы формирования луча [35]: *прямая оценка весовых коэффициентов формирования луча с помощью DNN*. В рамках этого подхода нейронные сети обучаются непосредственно определять и оценивать весовые коэффициенты многоканальных фильтров, используемых при формировании луча. Оптимизация этих коэффициентов в процессе обучения позволяет DNN улучшать качество речевого сигнала без необходимости предварительной оценки статистических свойств шума или чистого сигнала [14]. Среди передовых решений выделяются полностью нейросетевые системы формирования луча, в которых методы глубокого обучения интегрированы непосредственно в процесс обработки, что позволяет оптимизировать пространственную фильтрацию и подавление шума в рамках единой сквозной (end-to-end) архитектуры [7]; *прямая обработка многоканальных данных с помощью DNN*. По мере развития области исследований фокус сместился на системы сквозного обучения, работающие непосредственно с

необработанным аудиосигналом (волновой формой) и обучающиеся преобразовывать зашумленные входные сигналы в чистые выходные. В таких подходах сети глубокого обучения применяются непосредственно для обработки многоканальных данных, минуя традиционную структуру систем формирования луча.

DNN улучшают качество речи путем обработки необработанных многоканальных сигналов; часто при этом используются архитектуры сквозного обучения, позволяющие обрабатывать аудиосигналы без явного выделения признаков и эффективно учитывать временные зависимости [36].

Существуют также гибридные методы, сочетающие в себе черты вышеупомянутых подходов. Например, в некоторых методах DNN могут использоваться как для оценки статистических свойств, так и для оптимизации весовых коэффициентов фильтров, либо же традиционные методы формирования луча могут быть интегрированы в систему сквозного обучения. Эти достижения трансформировали традиционные подходы, обеспечив появление более совершенных, точных и эффективных методов обработки звука в сложных акустических условиях. Кроме того, для обработки сигналов с микрофонных решеток стали применяться модели, использующие механизм внимания (например, трансформеры); это позволяет лучше учитывать взаимосвязи между различными каналами и эффективнее фокусироваться на значимых источниках звука [36]. В последнее время также появились решения, основанные на диффузионных моделях и генеративных подходах, использующих возможности генеративных систем для улучшения качества обработки речи [8]. К числу актуальных тенденций относятся и мультимодальные подходы, сочетающие аудиоданные с визуальной или иной сенсорной информацией, что способствует повышению качества обработки речи и эффективности выделения целевого сигнала [37].

При многоканальной обработке речи ключевой проблемой является формирование обучающей выборки. В существующих методах многоканальные сигналы чаще всего создаются путем имитационного моделирования: многоканальные импульсные характеристики помещения (RIR) генерируются с использованием метода зеркальных источников. Одноканальные чистые сигналы источников берутся из различных открытых наборов данных, а шум обычно заимствуется из специализированных одноканальных датасетов или генерируется программно (например, в виде белого или пространственно-диффузного шума). Затем эти чистые одноканальные сигналы подвергаются свертке с импульсными

характеристиками помещения для получения многоканальных сигналов, которые впоследствии объединяются с шумом, образуя наблюдаемые сигналы. Однако такой подход к генерации многоканальных сигналов, основанный на моделировании, сопряжен с рядом трудностей.

Модель мнимых источников (image model), используемая для генерации импульсных характеристик помещения (RIR), ограничена представлением стандартных прямоугольных комнат [38]. Однако реальные акустические условия значительно сложнее: форма помещений может быть нестандартной, а внутри них могут находиться различные препятствия с неодинаковыми коэффициентами отражения. Вследствие этого модели мнимых источников могут неточно воспроизводить условия, характерные для реальной обстановки.

• Моделирование многоканальных сигналов путем свертки чистого сигнала источника с RIR не вполне соответствует действительности. Процесс свертки обычно предполагает неизменность акустического канала в течение определенного времени, тогда как в реальных условиях источники могут быть направленными и подвижными, а характеристики акустического канала часто меняются со временем.

• Моделирование многоканальных сигналов с учетом аддитивного шума может неточно отражать реальные ситуации, например, сценарий «коктейльной вечеринки». Это может быть одной из причин, по которым методы многоканальной обработки речи для улучшения ее качества пока не оправдывают ожиданий и не достигли такого же уровня успеха в реальных приложениях, как их одноканальные аналоги.

В задачах многоканального улучшения речи с использованием методов глубокого обучения вопрос о выборе целевого сигнала до конца не решен. На практике выбор целевого сигнала обычно сводится к двум вариантам: 1) чистый сигнал источника (сигнал прямого пути); и 2) сигнал прямого пути плюс ранние отражения, регистрируемые опорным микрофоном. Теоретически, если бы удалось полностью восстановить чистый сигнал источника, можно было бы обеспечить оптимальное качество обработки речи. Однако в реальных помещениях с реверберацией энергия сигнала прямого пути составляет лишь небольшую долю наблюдаемого сигнала, что делает полное восстановление этого сигнала крайне сложной задачей.

Более того, не все отражения негативно сказываются на восприятии речи. Принято считать, что ранние отражения могут способствовать улучшению качества и разборчивости речи (сигналы, приходящие в течение первых 20 мс,

в основном интегрируются слуховой системой и, как правило, увеличивают мощность полезного сигнала). Напротив, поздние отражения из-за своей некогерентности и большой длительности, превышающей время слуховой интеграции, могут ухудшать качество и разборчивость речи. Таким образом, целью часто становится оценка сигнала прямого пути вместе с ранними отражениями. В существующих методах к ранним отражениям обычно относят те, которые возникают в интервале менее 50 мс после прихода прямого звука. Этот порог основан на ранних психоакустических исследованиях и эмпирических данных. Однако влияние различных составляющих реверберационного сигнала на восприятие речи может варьироваться в зависимости от конкретных задач и условий. В связи с этим определение оптимального целевого сигнала и разработка наиболее эффективной модели обучения, опирающейся на данные, для многоканальной обработки речи остаются важными направлениями исследований.

Хотя за последнее десятилетие методы глубокого обучения существенно продвинули область обработки речи и аудиосигналов, они зачастую не учитывают фундаментальные физические аспекты решаемой задачи. Кроме того, методы, основанные исключительно на глубоком обучении, нередко характеризуются непрозрачностью и отсутствием интерпретируемости. Перспективным подходом к преодолению этих ограничений является включение в модели априорных знаний о данных и лежащих в их основе физических принципах [39]. Это предполагает сочетание глубокого обучения с традиционными методами обработки сигналов для создания физически обоснованных моделей, обеспечивающих лучшую обобщающую способность в различных условиях, средах и конфигурациях микрофонных решеток. Кроме того, при разработке моделей глубокого обучения в функции потерь следует учитывать метрики восприятия, поскольку на восприятие человеком влияют не только физические факторы и архитектура алгоритмов.

Основные достижения и ключевые факторы исследований

Общедоступные наборы данных и организованная конкуренция играют решающую роль в развитии исследований в области микрофонных решеток и многоканального улучшения качества речи. Эти ресурсы обеспечивают стандартизированные критерии оценки (benchmark), позволяя исследователям разрабатывать, тестировать и сравнивать алгоритмы в сопоставимых условиях. Одним из примеров в этой сфере является серия наборов данных CHiME (Computational Hearing in Multisource Environments — «Вычислительный слух в условиях с несколькими источниками звука»),

которая широко используется для оценки алгоритмов улучшения и распознавания речи в зашумленной среде [40]. С момента запуска в 2011 году было проведено семь этапов (выпусков) CHiME, благодаря чему проект стал важной платформой для развития надежных технологий обработки речи. Корпус AMI (Augmented Multi-party Interaction — «Расширенное многостороннее взаимодействие») содержит записи совещаний, сделанные с использованием нескольких микрофонных решеток [41]. Набор данных MIRaGe включает многоканальные записи, сделанные в акустической лаборатории с регулируемым временем реверберации [43]. Набор данных DIRHA (Distant-speech Interaction for Robust Home Applications — «Взаимодействие с использованием речи на расстоянии для надежных домашних систем») предлагает записи, сделанные в реальных домашних условиях с применением нескольких микрофонных решеток.

Эти наборы данных и конкуренция не только облегчают сравнительный анализ новых алгоритмов, но и способствуют воспроизводимости и прозрачности научных исследований. Однако, несмотря на значительный вклад этих ресурсов в данную область, по-прежнему ощущается нехватка стандартного, авторитетного и общепризнанного набора данных, подобного ImageNet в сфере обработки изображений. Появление новых наборов данных откроют дополнительные возможности для инноваций в области технологий микрофонных решеток и сфер их применения. Кроме того, во многих исследованиях сообщается об улучшении определенных показателей, которое может носить случайный характер, а методики проведения экспериментов зачастую описываются недостаточно подробно. Это подчеркивает важность открытой публикации данных и программного кода для повышения воспроизводимости результатов, а также для обеспечения возможности их проверки и дальнейшего использования в качестве основы для новых исследований.

Показатели эффективности играют ключевую роль в оценке и сравнении методов обработки сигналов с использованием микрофонных решеток. Ранние метрики зачастую были упрощенными и фокусировались преимущественно на улучшении отношения сигнал/шум (SNR) или степени подавления шума. Впоследствии были разработаны более комплексные показатели, лучше отражающие эффективность работы систем в реальных условиях; к ним относятся, например, STOI (Short-Time Objective Intelligibility) [44], оценивающий разборчивость речи в присутствии шума или искажений, а также PESQ (Perceptual Evaluation of Speech Quality) и POLQA (Perceptual

Objective Listening Quality Assessment), позволяющие оценить общее качество обработанной речи с учетом таких факторов, как четкость и естественность звучания.

Эти метрики способствовали унификации исследований, обеспечив общие критерии для оценки и сравнения различных алгоритмов. Несмотря на достигнутый прогресс, сохраняется ряд проблем.

Хотя данные метрики ценны для установления стандартов оценки, показатели, разработанные для одноканальных систем, могут не в полной мере учитывать сложности пространственной обработки звука и многоканальных сред. Кроме того, у этих метрик есть ограничения: например, улучшение показателя STOI не всегда коррелирует с лучшими результатами в ходе субъективных тестов с участием слушателей [45]. Некоторые исследования указывают на целесообразность разработки метрик качества и разборчивости, не требующих эталонного сигнала (non-intrusive measures) — хотя пока значительных успехов в этой области достичь не удалось, — а также метрик, пригодных для оптимизации и проверки в процессе работы системы (in-the-loop), поскольку большинство современных показателей используются преимущественно для финальной верификации [46]. Постоянное совершенствование и разработка новых метрик будут играть ключевую роль в точной оценке и дальнейшем развитии технологий микрофонных решеток.

Области применения

Микрофонные решетки стали неотъемлемой частью многих систем. В системах видеоконференцсвязи они играют решающую роль в обеспечении четкой и разборчивой передачи речи. По сравнению с использованием микрофона на гибком держателе, расположенного вблизи говорящего, применение технологии формирования луча (beamforming) в микрофонных решетках обеспечивает значительный выигрыш за счет направленности и более эффективное подавление шума, что улучшает общее качество восприятия звука участниками. Микрофонные решетки также способны отслеживать нескольких говорящих и динамически переключать фокус, что делает их идеальным решением для конференц-залов, где участники могут перемещаться или говорить из разных точек помещения.

Технологии микрофонных решеток также открывают новые возможности для слуховых аппаратов [47]. Учитывая жесткие ограничения по габаритам и энергопотреблению, характерные для таких устройств, компактные микрофонные решетки обладают огромным потенциалом. Современные слуховые аппараты, оснащенные микрофонными решетками и использующие

методы глубокого обучения, эффективнее различают речь и фоновый шум, тем самым облегчая пользователю восприятие разговора в шумной обстановке. Устройства для «умного дома», такие как голосовые помощники и умные колонки, активно используют микрофонные решетки для захвата голоса на расстоянии (far-field voice capture), что позволяет им распознавать голосовые команды и реагировать на них.

К числу новых областей применения относятся автономные транспортные средства, а также системы виртуальной (VR) и дополненной (AR) реальности. Микрофонные массивы преимущественно используются для совершенствования автомобильных голосовых помощников и систем взаимодействия, обеспечивая беспрепятственное общение между водителем и бортовым помощником. Такие массивы позволяют точно распознавать голосовые команды, что делает возможным эффективное управление навигацией, мультимедийными системами и другими функциями автомобиля. В приложениях виртуальной (VR) и дополненной (AR) реальности микрофонные массивы обеспечивают высококачественную передачу пространственного звука, записывая его с различных направлений, что критически важно для создания эффекта полного погружения пользователя. Ожидается, что эта перспективная технология вызовет значительный интерес при разработке будущих потребительских устройств.

Выводы

Технология микрофонных массивов достигла значительного прогресса и нашла широкое применение в различных областях. За последнее десятилетие стремительное развитие методов глубокого обучения сместило акцент в области микрофонных массивов и многоканальной обработки речи в сторону подходов, основанных на данных, что привлекло повышенное внимание исследователей. Несмотря на эти успехи, методы улучшения качества речи с использованием микрофонных массивов и глубокого обучения сталкиваются с рядом проблем, таких как необходимость формирования обширных наборов многоканальных обучающих данных, охватывающих разнообразные акустические условия. Существует большая потребность в создании стандартного набора данных, который позволил бы проводить простое и объективное сравнение различных алгоритмов. Еще одной сложной задачей является проектирование систем, учитывающих индивидуальные особенности слуха и условия эксплуатации у разных пользователей. Также необходимо решать вопросы, связанные с вычислительной сложностью и объемами данных, требуемыми для глубокого обучения, находя баланс между высокой

эффективностью и минимальными вычислительными затратами — особенно в системах реального времени, таких как мобильная связь и слуховые аппараты. Актуальной задачей остается и наделение микрофонных массивов способностью к полноценному пространственному слуху, позволяющему фокусироваться на одном или нескольких конкретных звуках подобно тому, как это делает слуховая система человека.

Литература

1. Garofolo J. S. TIMIT: Acoustic-Phonetic Continuous Speech Corpus. – Linguistic Data Consortium, 1993.

2. Gauvain J.-L., Lee C.-H. Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains // IEEE Transactions on Speech and Audio Processing. – 1994. – Vol. 2, No. 2. – P. 291–298.

3. Ghaemmaghami H., Baker B., Vogt R., Sridharan S. Noise robust voice activity detection using features extracted from the time-domain autocorrelation function // Proceedings of Interspeech. – 2010.

4. Godfrey J. J., Holliman E. C., McDaniel J. SWITCHBOARD: Telephone speech corpus for research and development // Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). – IEEE, 1992.

5. Good M. D., Gilkey R. H. Sound localization in noise: The effect of signal-to-noise ratio // The Journal of the Acoustical Society of America. – 1996. – Vol. 99. – P. 1108.

6. Graff D. An overview of Broadcast News corpora // Speech Communication. – 2002. – Vol. 37, No. 1. – P. 15–26.

7. Hain T., Burget L., Dines J., Garner P. N., Hannani A. E., Huijbregts M., Karafiat M., Lincoln M., Wan V. The AMIDA 2009 meeting transcription system // Proceedings of Interspeech. – 2010.

8. Hain T., Burget L., Dines J., Garner P., Grezl F., El Hannani A., Huijbregts M., Karafiat M., Lincoln M., Wan V. Transcribing meetings with the AMIDA systems // IEEE Transactions on Audio, Speech, and Language Processing. – 2012. – Vol. 20, No. 2. – P. 486–498.

9. Han K. J., Narayanan S. S. A robust stopping criterion for agglomerative hierarchical clustering in a speaker diarization system // Proceedings of Interspeech. – 2007.

10. Han K. J., Narayanan S. S. Novel inter-cluster distance measure combining GLR and ICR for improved agglomerative hierarchical speaker clustering //

Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). – 2008.

11. Hennecke M. D., Fink G. A. Towards acoustic self-localization of ad hoc smartphone arrays // 2011 Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA). – IEEE, 2011. – P. 127–132.

12. Hermansky H. Perceptual linear predictive (PLP) analysis of speech // The Journal of the Acoustical Society of America. – 1990. – Vol. 87, No. 4. – P. 1738–1752.

13. Himawan I., McCowan I., Lincoln M. Microphone array beamforming approach to blind speech separation // Machine Learning for Multimodal Interaction. – Springer, 2008. – P. 295–305.

14. Hori T., Hori C., Minami Y., Nakamura A. Efficient WFST-based one-pass decoding with on-the-fly hypothesis rescoring in extremely large vocabulary continuous speech recognition // IEEE Transactions on Audio, Speech, and Language Processing. – 2007. – Vol. 15, No. 4. – P. 1352–1365.

15. Hori T., Araki S., Yoshioka T., Fujimoto M., Watanabe S., Oba T., Ogawa A., Otsuka K., Mikami D. et al. Real-time meeting recognition and understanding using distant microphones and omni-directional camera // Proceedings of the Spoken Language Technology Workshop (SLT). – IEEE, 2010. – P. 424–429.

16. Huijbregts M. SHoUT speech recognition toolkit. – 2006. – http://shout-toolkit.sourceforge.net/ (murojaat sanasi: 2013-yil fevral).

17. Huijbregts M., van Leeuwen D. Large scale speaker diarization for long recordings and small collections // IEEE Transactions on Audio, Speech, and Language Processing. – 2012. – Vol. 20, No. 2. – P. 404–413.

18. Huijbregts M., van Leeuwen D. A., Wooters C. Speaker diarization error analysis using oracle components // IEEE Transactions on Audio, Speech, and Language Processing. – 2012. – Vol. 20, No. 2. – P. 393–403.

19. Huijbregts M. A. H. Segmentation, Diarization and Speech Transcription: Surprise Data Unraveled. – PhD thesis. – Centre for Telematics and Information, University of Twente, 2008.

20. Wang H., Ma Y., Zheng Q., Cao K., Lu Y., Xie H. Review of recent development of MEMS speakers // Micromachines. – 2021. – Vol. 12, No. 10. – Article 1257.

21. Gemelli A., Tambussi M., Fusetto S., Aprile A., Moisello E., Bonizzoni E., Malcovati P. Recent trends in structures and interfaces of MEMS transducers for audio applications: A review // Micromachines. – 2023. – Vol. 14. – Article 847.

22. Rusconi A., Costantini S., Prelini C. Micro speakers // Silicon Sensors and Actuators. – 2022. – P. 651.

23. Garud M., Pratap R. MEMS audio speakers // Journal of Micromechanics and Microengineering. – 2024. – Vol. 34. – Article 013001.

24. USound. Achlelous UT-P2020: Datasheet. – February 2021. – Revised June 2021.

25. Huber J. E., Fleck N. A., Ashby M. F. The selection of mechanical actuators based on performance indices // Proceedings of the Royal Society of London. Series A. – 1997. – Vol. 453. – P. 2185.

26. Harradine M. A., Birch T. S., Stevens J. C., Shearwood C. A micro-machined loudspeaker for the hearing impaired // Technical Digest of Papers of the Transducers'97 Conference. – 1997. – P. 429.

27. Cheng M.-C., Huang W.-S., Huang S. R.-S. A silicon microspeaker for hearing instruments // Journal of Micromechanics and Microengineering. – 2004. – Vol. 14. – P. 859.

28. Je S.-S., Rivas F., Diaz R. E., Kwon J., Kim J., Bakkaloglu B., Kiaei S., Chae J. A compact and low-cost MEMS loudspeaker for digital hearing aids // IEEE Transactions on Biomedical Circuits and Systems. – 2009. – Vol. 3. – P. 348.

29. Chen Y. C., Cheng Y. T. A low-power milliwatt electromagnetic microspeaker using a PDMS membrane for hearing aids applications // Proceedings of the 24th International Conference on Micro Electro Mechanical Systems (MEMS). – 2011. – P. 1213.

30. Shahosseini I., Lefeuvre E., Moulin J., Martinicic E., Woytasik M., Pillonnet G., Lemarquand G. Planar microcoil optimization of MEMS electrodynamic microspeakers // IEEE Transactions on Magnetics. – 2013. – Vol. 49, No. 8. – P. 4843.

31. Majlis B. Y., Sugandi G., Noor M. M. Compact electrodynamic MEMS-speaker // Proceedings of the Semiconductor Technology International Conference. – 2017.

32. Liechti R., Durand S., Hilt T., Casset F., Dieppedale C., Verdot T., Colin M. A piezoelectric MEMS loudspeaker lumped and FEM models // 22nd International Conference on Thermal, Mechanical and Multi-Physics Simulation and Experiments in Microelectronics and Microsystems (EuroSimE). – 2021.

33. Klasco M., Ricci M., Shaw N. An innovative MEMS piezo speaker // The Journal of the Acoustical Society of America. – 2021. – Vol. 149. – P. A94.

34. Anand J. J., Chithra R. B., Kumar S. Recent developments in MEMS speaker technology // Journal of VLSI Design Tools and Technology. – 2021. – Vol. 11. – P. 29.

35. Stoppel F., Fankhänel J., Giese T., Eisermann C., Pieper I., Kaden D., Castellanos L., Grünzig S. Highly miniaturized in-ear MEMS loudspeaker featuring high SPL // Proceedings of the 22nd International Conference on Solid-State Sensors, Actuators and Microsystems (Transducers). – Kyoto, Japan, 2023. – P. 865–868.

36. Xia Y., Balmes C., Barrera-Figueroa S., Jeong C.-H. Multi-channel calibration of microphone arrays based on the array related transfer function // 2026. – Vol. 254.

37. Fangeta S., Casseta F., Dejaegera R., Maireb F., Deslogesa B., Deutzerc J., Morissonc R., Bohardc Y., Larocheb B., Escatob J., Leclered Q. Low voltage MEMS digital loudspeaker array based on thin-film PZT actuators // Physics Procedia. – 2015.

38. Piao W., Yuan Y., Lin H. A Digital Image Denoising Algorithm Based on Gaussian Filtering and Bilateral Filtering // ITM Web of Conferences. – 2018. – Vol. 17.

39. You J., Chen Z., Song X., Shi Y., Chen Z. Robust multi-source DOA tracking using Gaussian kernel smoothed MeMBer filter for generalized coprime arrays // Signal Processing. – 2026. – Vol. 240.

40. Liu C., Zhao H. Efficient DOA Estimation Method Using Bias-Compensated Adaptive Filtering // IEEE Transactions on Vehicular Technology. – 2026. – Vol. 69, No. 11.

41. Salama A. A. Direction of Arrival Estimation: A Tutorial Survey of Classical and Modern Methods. – 2025.

42. Стринкевич Э. А., Чекан В. Л. Слуховая система человека: учебно-методическое пособие. – Минск: БелМАПО, 2016.

43. Li Z., Wang J., Li X., Li W., Luo L., Li L., Hong Q. Speaker Diarization with Overlapping Community Detection Using Graph Attention Networks and Label Propagation Algorithm. – 2025.

44. Coenell S. Overlapped Speech Detection and Speaker Counting Using Distant Microphone Arrays // Computer Speech & Language. – 2022. – Vol. 72.

45. Sun Z., Zhang L., Wang Q., Zhou P., Xie L. Towards Robust Overlapping Speech Detection: A Speaker-Aware Progressive Approach Using WavLM. – 2025.

46. Raghavendra S., Chun H., Lee S., Chen F., Martin B. A., Tan C.-T. Cross-Frequency Coupling in Cortical Processing of Speech. – Scottish Event Campus, Glasgow, UK, 11–15 July 2022.

47. Yan L., Huan J., Gao H., Zhang W., Ma J. Processing of Speech Signals Based on MATLAB & FDAtool. – Xi'an Eurasia University, Xi'an, Shaanxi, China, 2024.