

RU

Цифровые исследования звучащей речи:
история, методология, современные инструменты

Гончарова О. В.

Аннотация. Цель исследования – установить взаимосвязи между историческими этапами развития технологий обработки речевых данных и формированием современной методологической базы инструментальной фонетики. В статье представлена эволюция становления инструментальных подходов к изучению звучащей речи – от первых механических устройств XVIII века до современных нейросетевых архитектур, проанализированы программно-технические решения современности – от универсальных платформ акустического анализа до специализированных систем автоматического выравнивания речевых данных. Исследование включает систематизацию методологических принципов каждого исторического периода с выявлением их концептуальных ограничений и потенциала для решения актуальных лингвистических задач. Научная новизна заключается в периодизации развития инструментальной фонетики, основанной на взаимодействии технологических возможностей и методологических концепций каждого этапа. В результате были выявлены три типа концептуальных разрывов (технологический, семантический, когнитивный), препятствующих эффективной интеграции современных цифровых технологий с традиционными лингвистическими категориями. Обоснована необходимость создания гибридных аналитических платформ, способных преодолеть фрагментацию между количественными параметрами автоматической обработки и качественными характеристиками фонологического описания.

EN

Digital studies of spoken speech: history, methodology, modern tools

O. V. Goncharova

Abstract. The aim of the research is to establish the interrelationships between the historical stages of development of speech data processing technologies and the formation of the modern methodological base of instrumental phonetics. The paper presents the evolution of the formation of instrumental approaches to the study of spoken speech – from the first mechanical devices of the 18th century to modern neural network architectures, and analyzes modern software and hardware solutions – from universal platforms for acoustic analysis to specialized systems for automatic alignment of speech data. The study includes a systematization of the methodological principles of each historical period, identifying their conceptual limitations and potential for solving current linguistic problems. The scientific novelty lies in the periodization of the development of instrumental phonetics, based on the interaction of technological capabilities and methodological concepts of each stage. As a result, three types of conceptual gaps (technological, semantic, cognitive) were identified, hindering the effective integration of modern digital technologies with traditional linguistic categories. The necessity of creating hybrid analytical platforms capable of overcoming the fragmentation between quantitative parameters of automatic processing and qualitative characteristics of phonological description is substantiated.

Введение

Актуальность настоящего исследования определяется не только стремительным внедрением технологий искусственного интеллекта в прикладные сферы, но и трансформацией самих парадигм лингвистического анализа, связанной с переходом от описательных моделей к интерактивным и самообучающимся системам обработки речи. Современное языкознание сталкивается с задачей осмысления огромных массивов акустических и мультимодальных данных, генерируемых голосовыми ассистентами, системами автоматического распознавания и синтеза речи. В таких условиях классические методики качественного отбора и интерпретации речевых феноменов требуют пересмотра и дополнения с учетом возможностей статистического моделирования, глубоких нейронных сетей и гибридных подходов, объединяющих эксперименты с корпусными

данными и визуализацией артикуляторных процессов. Нарастающая интеграция инструментальных фонетических методов и алгоритмов глубокого обучения создает методологический вызов: необходимо выработать четкие критерии выбора технических средств и аналитических процедур для решения разнородных лингвистических задач – от оценки динамики формантных структур в экспериментальной фонетике до построения адаптивных образовательных платформ, способных в реальном времени отслеживать и корректировать произносительные особенности. При этом возникает принципиальная потребность в установлении соответствий между параметрами цифровой обработки и традиционными языковыми категориями – фонемами, просодическими единицами, социолингвистическими переменными. Исследование истории, методологии и современных инструментов анализа звучащей речи представляет собой необходимую предпосылку для формирования универсальных, масштабируемых и когнитивно интерпретируемых лингвистических платформ. Выяснение преемственности технологических этапов и выявление «узких мест» в интеграции разных уровней анализа позволит не только оптимизировать экспериментальные процедуры, но и обосновать разработку новых теоретико-методологических моделей, адекватных вызовам цифровой эпохи.

Задачи исследования: 1) выявить основные этапы становления инструментальных методов анализа звучащей речи с учетом технологических принципов и теоретических оснований каждого периода; 2) проанализировать эволюцию методологических подходов в цифровой обработке речевого сигнала; 3) систематизировать современные программно-технические решения для анализа звучащей речи с определением их функциональных возможностей и области применения; 4) выявить концептуальные разрывы между количественными параметрами цифрового анализа и лингвистическими категориями фонетического описания.

Методы исследования: в работе применяется сравнительно-сопоставительный метод для изучения развития инструментальных методов фонетики и выявления общих принципов и различий в функциональных возможностях современных программных платформ цифрового анализа речи (Praat, ELAN, Montreal Forced Aligner, WebMAUS, Articulate Assistant Advanced), а также корпусный анализ с использованием цифровых инструментов для систематизации и классификации источников.

Теоретическую базу настоящего исследования составляют труды отечественных и зарубежных ученых, посвященные проблемам теории фонемы и вопросам фонологии (Щерба, 1931; Трубецкой, 2000; Бондарко, 1998), акустической фонетики (Фант, 1964; Stevens, 1998), цифровой обработки сигналов (Oppenheim, Schaffer, Yuen, 1978), формантного анализа гласных (Peterson, Barney, 1952). Также были использованы труды основателей экспериментальной фонетики, в которых рассматриваются первые механические и аналоговые методы анализа речевого сигнала (Tillmann, 2006), ключевые публикации по цифровой обработке и статистическому моделированию речи (Juang, Rabiner, 2005) и обзоры по синтезу речи (Taylor, 2009), а также публикации, посвященные алгоритмам автоматического выравнивания звуковых и текстовых данных: Montreal Forced Aligner (McAuliffe, Socolof, Mihuc et al., 2017) и WebMAUS (Kisler, Reichel, Schiel, 2017).

Для отражения современного состояния исследований были проанализированы технические описания и пользовательские руководства программ Praat (Boersma, Weenink, 2025) и ELAN (Max Planck Institute for Psycholinguistics), документация Speech Filing System (1998).

Теоретическая значимость исследования состоит в выстраивании целостной схемы, объединяющей этапы развития инструментальной фонетики и цифровой обработки речевого сигнала, позволяя по-новому осмыслить роль каждого этапа в формировании современных исследований речи, а также в преодолении дихотомии между технологической практикой и лингвистической теорией, показывая, что цифровые методы не просто упрощают сбор и обработку данных, но создают новые условия для формулирования научных вопросов и интерпретации понятий (например, фонема как статистический аттрактор в пространстве признаков).

Практическая ценность результатов исследования состоит в обеспечении методологического обоснования для разработки новых исследовательских протоколов, учитывающих специфику цифровой обработки речевых данных при сохранении связи с традиционными лингвистическими категориями, а также основы для принятия решений при выборе исследовательских инструментов, позволяя соотносить специфику решаемых задач с техническими возможностями и ограничениями различных методов анализа речи. Полученные результаты могут быть использованы в процессе преподавания таких дисциплин, как прикладная лингвистика (для иллюстрации эволюции инструментальных методов анализа речи и их интеграции с цифровыми технологиями), экспериментальная фонетика (в контексте систематизации исторических этапов и современных инструментов для акустического и артикуляторного анализа), компьютерная лингвистика и обработка естественного языка (Natural Language Processing, NLP) (при изучении алгоритмов автоматического распознавания и синтеза речи, включая нейросетевые модели), а также в междисциплинарных курсах по искусственному интеллекту в гуманитарных науках (для демонстрации концептуальных разрывов между технологическими и лингвистическими подходами и обоснования гибридных платформ). Кроме того, материалы исследования применимы в образовательных программах по речевым технологиям, способствуя формированию у студентов навыков работы как с корпусными данными, так и с инструментами, такими как Praat, ELAN и Montreal Forced Aligner.

Обсуждение и результаты

История инструментальных методов в фонетике неразрывно связана с развитием научного познания и технологического прогресса, уходя корнями в XVIII век. Первые значимые шаги в данной области были

предприняты в направлении механического воспроизведения человеческой речи, что заложило фундамент для последующих, более сложных исследований. Одной из первых задокументированных попыток механического синтеза речи стало исследование датского физиолога Кристиана Кратценштейна. В 1779 К. Кратценштейн предложил решение для конкурса на объяснение различий между пятью основными гласными звуками, создав систему резонансных трубок, которые, приводимые в действие пульсирующим воздушным потоком, могли воспроизводить звуки (Фланаган, 1968; Соломенник, 2013). Практически одновременно, но независимо от К. Кратценштейна, Вольфганг фон Кемпелен работал над собственной механической системой синтеза речи, представив в 1791 году «Акустико-механическую речевую машину». Устройство В. Кемпелена моделировало прохождение воздушной струи через речевой тракт человека, используя меха для подачи воздуха и управляемый резонатор. Согласные звуки, включая носовые, формировались с помощью четырех каналов, зажимаемых пальцами (Фланаган, 1968). По утверждению В. Кемпелена, машина могла производить 19 хорошо различимых согласных и короткие фразы на нескольких языках (Кейтер, 1985; Mattingly, 1974; Соломенник, 2013). Усовершенствованный вариант машины В. Кемпелена был создан в 1837 году английским физиком Чарльзом Уитстоном, который использовал резонаторы из кожи, конфигурацию которых можно было изменять вручную для получения различных звуков, похожих на речь (Фланаган, 1968).

Расцвет экспериментальной и инструментальной фонетики начался в конце XIX века, во многом благодаря усилиям аббата П. Ж. Руссело (Abbé Rousselot) (Rousselot, 1897) в Париже. Его диссертация 1891 года и труд «Principes de phonétique expérimentale» (1897) стали основополагающими в данной области. П. Ж. Руссело основал лабораторию в Коллеж де Франс в 1898 году, что послужило примером для создания аналогичных учреждений по всему миру (Galazzi, 2014).

Также в данный период были разработаны и активно применялись первые инструменты для визуализации и анализа речевых процессов:

- фоноавтограф, изобретенный Э. Л. Скоттом де Мартенвилем в 1857 году, стал одним из первых приборов для визуализации звуковых колебаний. Устройство фиксировало колебания мембраны, вызываемые звуком, на закопченной стенке вращающегося барабана, позволяя «увидеть» волну речи на бумаге (Gafni, 2025);
- фонограф, который представлял собой первый аппарат для записи и воспроизведения звука. Фонограф быстро нашел применение в фонетических экспериментах, и уже в 1878 году Ф. Дженкин и Дж. Юинг провели анализ гласных, выделяя форманты (Gafni, 2025);
- палатография, зародившаяся в начале XIX века с опытов Эразма Дарвина, в которых для изучения артикуляции гласных в полость рта помещались цилиндры из консервной жести. В 1872 году английский стоматолог О. Колз предложил прямую палатографию, при которой на нёбо наносилась мучная паста, а затем фиксировались следы контакта языка. В 1888 году Рихард Ленц обобщил накопленный опыт в своей работе «О физиологии и истории палатальных», дополнив прямое наблюдение косвенной техникой, например, приклеиванием к нёбу съемных пластин, что позволило получить более высокое разрешение данных. Благодаря введению метрической сетки на поверхности палатографа стало возможным не просто регистрировать факт контакта, но и измерять его точное положение вдоль передней-задней и медиальной-латеральной осей нёба (Tillmann, 2006, p. 374-389);
- кимограф, широко применявшийся в фонетических лабораториях с 1890-х по 1950-е годы, использовал датчики давления, соединенные с воздушными шлангами, для процарапывания следов на закопченной бумаге. Инструмент позволял отслеживать динамику артикуляции, включая изменения атмосферного давления, движения мышц и контур акустического сигнала, измерять длительность звуков, особенности дыхания и общий контур речевой волны, например, различать шумные и тоновые интервалы (Gósy, 2011).

С наступлением XX века и развитием электроники произошел фундаментальный сдвиг в подходах к синтезу и анализу речи. Механические системы постепенно уступали место электрическим аналогам, что открыло новые возможности для более точного моделирования речевых процессов.

Важным этапом в развитии экспериментальных фонетических исследований и синтеза речи стала разработка звукового спектрографа в Bell Labs в 1946 году (Соломенник, 2013). Первые образцы спектрографов, изначально использовавшиеся в военных целях для кодирования и декодирования связи, позволили разложить речь на спектральные составляющие и получать «звуковую спектрограмму». В послевоенное время, с коммерческим выпуском аппаратов типа Kay Sonagraph, спектрограф стал стандартным инструментом фонетических лабораторий, позволяющим визуализировать развитие формант звука во времени и проводить количественный анализ амплитудно-частотной структуры речи (Sound and Science).

С середины XX века фонетика претерпела существенные изменения, связанные с появлением и бурным развитием вычислительной техники (Таблица 1). К 1950-1960-м годам стало очевидно, что компьютер способен заменить любую аналоговую обработку сигнала, а к 1970-м годам в фонетических лабораториях появились крупные вычислительные комплексы, предназначенные для анализа оцифрованной речи и управления синтезаторами. Появление быстрого преобразования Фурье (Fast Fourier Transform, FFT) и алгоритмов линейного предсказания (Linear Predictive Coding, LPC) в 1960-1970-е годы стало революционным прорывом, позволив анализировать спектры речи в режиме реального времени. В 1980-е годы наступил век статистического моделирования, который кардинально изменил подходы к распознаванию и синтезу речи. Основой стало внедрение скрытых марковских моделей (Hidden Markov Models, HMM) с многокомпонентными гауссовыми смещениями, что позволило формализовать задачу распознавания как оптимизацию вероятностного решения (Juang, Rabiner, 2005). Были разработаны методы расчета первых и вторых производных спектральных коэффициентов (дельта- и дельта-дельта-цепстра), а также языковые модели на основе N-грамм. HMM стали доминирующей парадигмой в автоматическом распознавании речи (ASR) на протяжении десятилетий (Juang, Rabiner, 2005).

Параллельно со статистическим моделированием возродился интерес к искусственным нейронным сетям (Artificial Neural Networks, ANN), которые, хотя и появились ранее, получили новый импульс развития с возникновением больших объемов данных и вычислительных мощностей. Так, развитие цифровой обработки сигналов и статистического моделирования привело к значительному прогрессу в двух ключевых областях: автоматическом распознавании речи (Automatic Speech Recognition, ASR) и синтезе речи (Text-to-Speech, TTS). С 1970-х годов ASR, целью которого является преобразование устной речи в текстовую форму, стало основой для голосовых интерфейсов, систем диктовки и автоматизации колл-центров (Juang, Rabiner, 2005). Синтез речи (TTS), направленный на генерацию искусственной речи из текста, также претерпел значительные изменения. С 1990-х годов TTS-системы стали широко использоваться в голосовых помощниках, аудиокнигах и системах оповещения (Taylor, 2009).

Таблица 1. Основные цифровые методы обработки речевого сигнала и их применение

Метод/Технология	Год/Период	Принцип работы	Применение	Эволюционное значение
Цифровая обработка сигналов (ЦОС)	С 1960-х гг.	Преобразование аналогового речевого сигнала в цифровую форму для компьютерной обработки	Основа для всех цифровых методов анализа и синтеза речи	Фундаментальный сдвиг в фонетике, повышение точности и скорости анализа
Линейное предсказательное кодирование (LPC)	1970-е гг.	Моделирование речевого тракта как линейной системы для предсказания сигнала	Сжатие речи, извлечение параметров речевого тракта	Эффективное представление речевого сигнала, снижение требований к вычислительным ресурсам
Автоматическое распознавание речи (ASR)	С 1970-х гг.	Преобразование устной речи в текстовую форму	Голосовые интерфейсы, диктовка, автоматизация колл-центров	Позволило машинам «понимать» человеческую речь, основа для голосовых помощников
Скрытые марковские модели (HMM)	С 1980-х гг.	Статистическое моделирование последовательностей речевых звуков с учетом изменчивости	Распознавание речи, сегментация речи	Доминирующая парадигма в ASR на протяжении десятилетий, эффективное управление вариативностью речи
Искусственные нейронные сети (ANN) и глубокое обучение	С 1980-х гг.	Имитация работы человеческого мозга для обучения на больших данных	Распознавание речи, синтез речи, понимание естественного языка	Повышение точности и производительности в ASR и TTS, новые возможности для анализа речи
Синтез речи (TTS)	С 1990-х гг.	Генерация искусственной речи из текста	Голосовые помощники, аудиокниги, системы оповещения	Позволило машинам «говорить», расширение доступности информации

Важным также представляется отметить, что с появлением персональных компьютеров (1980-1990-е годы) активно развивались специализированные программы для фонетического анализа, которые значительно расширили возможности исследователей. В 1987 году была представлена система Speech Filing System (SFS) – универсальный формат файлов и пакет программ для хранения и анализа речевых данных (SFS, 1998). SFS предоставила стандартизированный подход к работе с речевыми данными, что способствовало обмену информацией и унификации исследований. Созданный в 1992 году П. Боерсмой и Д. Веенинком (Boersma, Weenink, 2025), Praat является одним из наиболее широко используемых программных инструментов для акустического анализа речи. Написанный на языке C++ и обеспечивающий кроссплатформенную работу, Praat предоставляет исследователю полный спектр возможностей для акустического анализа речи, включая: построение спектрограмм, визуализирующих эволюцию частотного состава звукового сигнала во времени; извлечение и графическую интерпретацию основной частоты (F0); автоматическую и ручную сегментацию формантных пиков с последующим количественным описанием качества гласных.

Для работы с мультимедийными данными, когда необходимо синхронно анализировать аудио- и видеозаписи, традиционно применяется ELAN, разработанный Институтом психолингвистики Макса Планка (Max Planck Institute for Psycholinguistics, 2025). Главным достоинством ELAN является возможность создавать неограниченное число слоев, структурированных в виде иерархии, что позволяет одновременно отображать фонетические транскрипции, морфологическую разметку, переводы и произвольные комментарии в тесной привязке ко временным меткам медиафайлов. Помимо прочего, ELAN обеспечивает совместимость с Praat и поддерживает подключение внешних инструментов, что делает его незаменимым для мультимодальных исследований.

Также современные системы способны автоматически выравнивать текстовый транскрипт со звуковой записью, что значительно ускоряет процесс аннотирования больших корпусов речи. Одним из таких инструментов является Montreal Forced Aligner (MFA) – инструмент с открытым исходным кодом, разработанный коллективом Университета Макгилла для решения задачи автоматической привязки речевых записей к их текстовым транскрипциям (McAuliffe, Socolof, Mihuc et al., 2017). MFA выравнивает транскрипт с соответствующим аудиофайлом на уровне слов и фонем, что обеспечивает высокую точность сегментации. Еще один известный сервис – WebMAUS: веб-приложение для автоматической сегментации речи (Kisler, Reichel, Schiel, 2017). На примере структуры TextGrid (Рисунок 1) видно, что в ходе обработки реализуется трехуровневая иерархия аннотаций, отражающая последовательное преобразование исходного орфографического текста в приближенную к фонетической сегментацию.

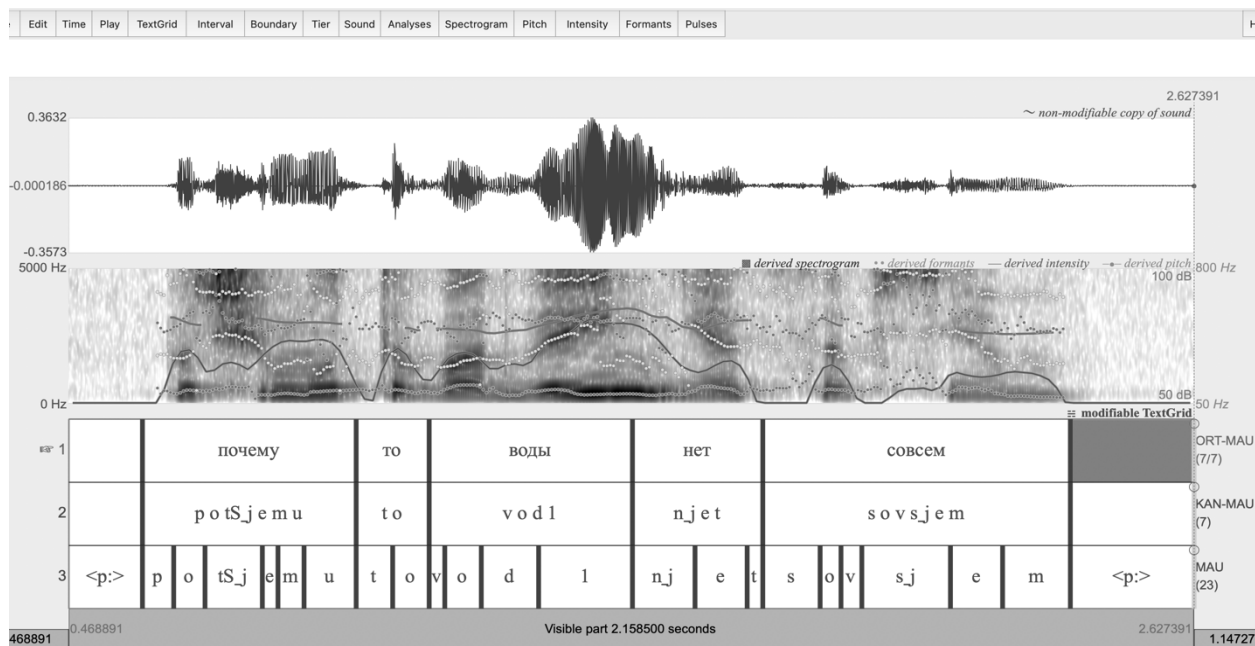


Рисунок 1. Результат выравнивания текста со звуковым файлом, выполненного WebMAUS

Первый уровень ORT-MAU содержит орфографические метки слов («почему», «то», «воды», «нет», «совсем») с указанием временных границ, интервалы пауз помечаются тегом <p>. Второй уровень KAN-MAU отражает упрощенную транскрипцию, где, например, (tS_j) соответствует аффрикате (tɕ), подчеркивание обозначает палатализацию. Третий уровень MAU фиксирует временные границы каждого сегмента, однако, несмотря на детализацию, MAU-уровень не представляет собой фонемной транскрипции в строгом лингвистическом смысле, так как для обозначения звуков используются адаптированные условные символы и не учитываются контекстуальные фонетические изменения.

Помимо акустических методов, для изучения артикуляции речи активно применяются инструментальные методы, позволяющие визуализировать движения речевого аппарата в процессе производства речи. Наиболее распространенным методом является ультразвуковая томография языка. Специальный ультразвуковой датчик получает данные о положении кончика языка, а программное обеспечение строит динамичное изображение (Stone, 2005). Для записи и анализа таких данных часто используется Articulate Assistant Advanced (AAA), разработанный компанией Articulate Instruments Ltd и функционирующий как надстройка над универсальным форматом SFS (Articulate Instruments).

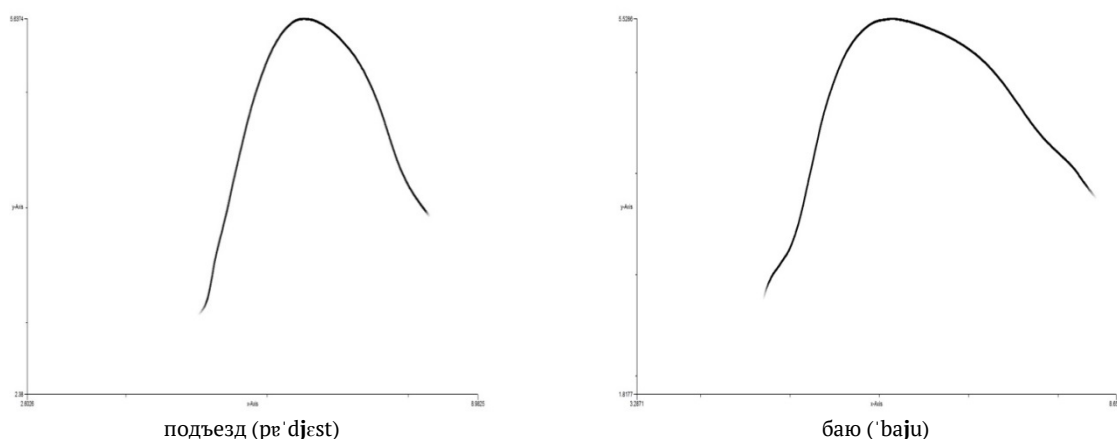


Рисунок 2. Проекция движения языка во время произнесения “j”, выполненная в программе Articulate Assistant Advanced

Истоки данной системы восходят к ранним версиям Articulate Assistant, ориентированным преимущественно на электропалатографию. С течением времени функциональность AAA была значительно расширена, включив поддержку ультразвуковой визуализации языка, захват видеопотока лица и губ, а также данные электромагнитной артикулографии (ЕМА) и электропалатографии (EPG). Архитектурно AAA реализован как модульная система рабочих пространств (workspaces), каждое из которых ориентировано на специфические типы данных и операций: в ультразвуковом пространстве предоставляется интерфейс для последовательного просмотра и разметки кадров; в “spline” пространстве возможно автоматизированное извлечение и усреднение контуров

языка (Рисунок 2). Модульный дизайн программного обеспечения позволяет исследователю конфигурировать рабочее пространство в зависимости от используемых методов сбора данных и поставленных задач: от трехмерного моделирования траекторий артикуляторов до измерения расстояний, углов и скоростей их движения.

Таким образом, на основании проведенного анализа представляется возможным выделить четыре основных этапа развития инструментальной фонетики, каждый из которых характеризуется доминирующими технологическими принципами, методологическими подходами и ограничениями (Таблица 2):

- механический этап развития технологий синтеза речи, охватывающий период с XVIII по середину XIX века, характеризовался первыми попытками физического моделирования артикуляционного аппарата человека. В основе исследований данного периода лежала концепция речи как совокупности физических процессов, которые могут быть воспроизведены механическими средствами. Существенным недостатком механических систем являлось отсутствие инструментальных средств для количественного анализа речевых сигналов;

- аналогово-инструментальный этап (вторая половина XIX – первая половина XX века) отмечен методологическим сдвигом от синтетического подхода к аналитическому исследованию речевых процессов. Ключевой особенностью данного периода стало создание специализированного измерительного оборудования, обеспечившего переход от моделирования речи к ее объективному исследованию, что положило начало экспериментальной фонетике как самостоятельной научной дисциплине;

- электронно-аналоговый этап (1920-е – 1970-е годы) представляет качественно новую фазу развития речевых технологий, обусловленную интеграцией электронных систем в процессы анализа и синтеза речи, а также переход от традиционных одномерных временных представлений речевых данных к двумерным спектрографическим изображениям. Важным достижением периода стало формирование стандартизированных процедур акустического анализа, обеспечивших воспроизводимость и сопоставимость результатов исследований. Теоретическую основу электронно-аналогового этапа составляла концепция речи как сложного многопараметрического акустического феномена, требующего комплексного описания с учетом множественных характеристик;

- цифровой этап (с 1960-х годов по настоящее время) характеризуется фундаментальным изменением парадигмы исследований благодаря внедрению вычислительных методов обработки речевых данных. Этот период подразделяется на несколько подэтапов: ранний цифровой подэтап (1960-1980-е годы) с внедрением алгоритмов быстрого преобразования Фурье и линейного предсказания; статистический подэтап (1980-2000-е годы) с доминированием скрытых марковских моделей; нейросетевой подэтап (с 2000-х годов) с развитием глубоких архитектур машинного обучения. Эволюция цифрового этапа отражает переход от простых алгоритмов к вероятностным моделям и далее к системам искусственного интеллекта, способным к самообучению на больших массивах речевых данных.

Таблица 2. Сравнительный анализ методологических этапов исследования звучащей речи

Этап	Базовые принципы	Используемые аппараты/модели	Типы получаемых данных	Способы интерпретации результатов	Критерии оценки качества анализа
Механический (XVIII – середина XIX в.)	Моделирование артикуляционного аппарата как физического устройства; речь как сумма резонансных колебаний	Простейшие резонансные модели (системы трубок, меха)	Акустические сигналы (звуки резонансных трубок); механические паттерны	Качественное сопоставление выходных звуков с естественной речью; слуховая оценка	Субъективная различимость звуков; способность генерировать ограниченный набор гласных/согласных
Аналогово-инструментальный (вторая пол. XIX – первая пол. XX в.)	Объективная визуализация и измерение речевого сигнала; переход от синтеза к экспериментальному анализу	Аналоговые осциллографы, кинографы; метрическая сетка палатографии	Записи колебаний мембраны (фоноавтограф), спектрограммы (фонограф), траектории артикуляции (палатография, кимограф)	Анализ формы волны и спектра на бумаге; измерения длительности, амплитуды, положения контактов артикуляторов	Точность метрических измерений (длина, угол, время); повторяемость визуализаций; пространственное/временное разрешение
Электронно-аналоговый (1920-е – 1970-е гг.)	Преобразование временной формы сигнала в частотно-временной анализ; стандартизация процедур фонетического исследования	Спектрографы; аналоговые фильтры; ламповые усилители	Частотно-временные спектрограммы; амплитудно-частотные характеристики	Квантификация формантных траекторий; графическое сравнение спектров	Разрешение по частоте и времени; точность выделения формант; сопоставимость результатов между лабораториями
Цифровой (с 1960-х гг. – н. в.)	Вычислительная обработка, статистическое моделирование и обучение; речь как многомерный цифровой объект	FFT, LPC, HMM, n-граммные модели; глубокие нейронные сети	Дискретизированные сигналы; спектральные коэффициенты; вероятностные траектории; латентные представления нейросетей	Статистический анализ параметров (форманты, Mel-Frequency Cepstral Coefficients, MFCC); вероятностная сегментация и классификация; кластеризация латентных пространств	Метрики распознавания (Word Error Rate, WER, Phone Error Rate, PER); качество синтеза (Mean Opinion Score, MOS); точность выравнивания; устойчивость моделей при вариативности речи

Таким образом, представленная схема развития инструментальной фонетики демонстрирует не только технологический прогресс, но и эволюцию подходов к пониманию речи как объекта научного исследования. Каждый этап характеризуется специфическими возможностями и ограничениями, понимание которых является необходимым условием для выбора адекватных методологических решений в современных исследованиях звучащей речи.

Заключение

В заключение следует подчеркнуть, что настоящее исследование призвано не просто зафиксировать хронологию появления инструментальных методов, но и выявить закономерности трансформации самой парадигмы речевого анализа под влиянием технологических изменений. Проведенный анализ показал, что эволюция от механических устройств к современным нейросетевым системам представляет собой качественную трансформацию концептуальных оснований фонетики как научной области в рамках языкознания. Ключевой вывод состоит в том, что каждый технологический этап формировал собственную модель речевого сигнала. Механический этап создавал представление о речи как физическом процессе, доступном прямому моделированию. Появление электронного спектрографа кардинально изменило не только инструментарий, но и саму концепцию фонетического анализа, предложив многопараметрическое описание звуковых единиц через частотно-временные характеристики. Цифровая революция привнесла абстрактные математические модели – от алгоритмов быстрого преобразования Фурье до скрытых марковских цепей, – что существенно расширило аналитические возможности, но одновременно усложнило связь между измеряемыми параметрами и лингвистическими сущностями. Современный нейросетевой этап достигает высокой практической эффективности за счет латентных представлений, интерпретация которых в терминах традиционной фонетики остается проблематичной.

Результатом проведенного анализа является выявление трех взаимосвязанных концептуальных разрывов. Технологический разрыв проявляется в несовместимости методологических принципов между этапами развития, что требует постоянной адаптации исследовательских процедур. Семантический разрыв обнаруживается в нарастающей дистанции между количественными показателями приборов и качественными категориями лингвистического описания. Когнитивный разрыв выражается в том, что современные алгоритмы, обеспечивающие высокую точность распознавания и синтеза речи, оперируют параметрами, не имеющими прямых коррелятов в системе фонологических и просодических единиц.

Данные разрывы не являются простыми техническими недостатками, но отражают фундаментальную проблему современной инструментальной фонетики – необходимость выработки новой методологии, способной интегрировать достижения цифровых технологий с лингвистической теорией. Перспективным направлением представляется создание комплексных аналитических платформ, объединяющих различные уровни речевого анализа в едином концептуальном пространстве, что позволит преодолеть существующую фрагментацию данных и обеспечить когнитивно-ориентированную интерпретацию результатов автоматической обработки речи. В ответ на обозначенные вызовы автором (Гончарова, 2024, с. 24-25) инициирована разработка платформы PySound, задача которой заключается в объединении этапов автоматической транскрипции и сегментно-суперсегментного анализа, оперирующих как статистическими паттернами, так и лингвистическими категориями (предварительная апробация PySound состоялась на конференции «Международные Иванниковские чтения» (2024), где был представлен модуль автоматической транскрипции, а также на конференции «Фонетика сегодня» (2024), где демонстрировалось объединенное сегментно-суперсегментное пространство анализа).

Материалы исследования | Research materials

1. Articulate Instruments. <https://www.articulateinstruments.com>
2. Boersma P., Weenink D. Praat: doing phonetics by computer (Version 6.1.03) (Computer software). 2025. <http://www.praat.org/>
3. Max Planck Institute for Psycholinguistics. ELAN (EUDICO Linguistic Annotator) (Computer software). <https://archive.mpi.nl/tla/elan>
4. PySound. <https://github.com/brainteaser-ov/PySound>
5. Sound and Science. <https://soundandscience.net/>
6. Speech Filing System (SFS). 1998. <https://www.phon.ucl.ac.uk/resource/sfs/>
7. Rousselot P. J. Principes de phonétique expérimentale. P.: H. Welter, 1897.

Источники | References

1. Бондарко Л. В. Фонетика современного русского языка. СПб.: Изд-во С.-Петербург. ун-та, 1998.
2. Гончарова О. В. Pysound – цифровой сервис обработки и анализа звучащей речи // Фонетика сегодня: тезисы докладов IX Международной научной конференции (г. Москва, 5-7 декабря 2024 г.). М., 2024.
3. Кейтер Дж. Компьютеры – синтезаторы речи. М.: Мир, 1985.

4. Соломенник А. И. Технология синтеза речи: история и методология исследований // Вестник Московского университета. Серия 9: Филология. 2013. № 6.
5. Трубецкой Н. С. Основы фонологии. М.: Аспект пресс, 2000.
6. Фант Г. Акустическая теория речеобразования / пер. с англ. Л. А. Варшавского, В. И. Медведева; под ред. В. С. Григорьева. М.: Наука, 1964.
7. Фланаган Дж. Л. Анализ, синтез и восприятие речи / пер. с англ.; под ред. А. А. Пирогова. М.: Связь, 1968.
8. Щерба Л. В. О тройном аспекте языковых явлений и об эксперименте в языкознании // Известия Отделения русского языка и словесности Академии наук СССР. 1931. № 1.
9. Gafni Ch. Phonetics and Phonology: An Introduction to the Science of Speech. 2025. https://www.researchgate.net/publication/388791051_Phonetics_and_Phonology_An_Introduction_to_the_Science_of_Speech
10. Galazzi E. Pierre Jean Rousselot: la phonétique expérimentale au service de l'homme. Dossiers d'HEL // Linguistiques d'intervention. Des usages socio-politiques des savoirs sur le langage et les langues. 2014. <https://shs.hal.science/halshs-01115159v1/document>
11. Gósy M. From stomatoscopy to BEA: the history of Hungarian experimental phonetics // Proceedings of 17th International Congress of Phonetic Sciences (Hong Kong, City University of Hong Kong). Hong Kong, 2011.
12. Juang B., Rabiner L. Automatic Speech Recognition – A Brief History of the Technology Development. 2005. https://www.researchgate.net/publication/249888949_Automatic_Speech_Recognition_-_A_Brief_History_of_the_Technology_Development
13. Kisler T., Reichel U. D., Schiel F. Multilingual processing of speech via web services // Computer Speech & Language. 2017. Vol. 45.
14. Mattingly I. G. Speech Synthesis for Phonetic and Phonological Models // Current Trends in Linguistics / ed. by T. S. Sebeok. The Hague: Mouton, 1974. Vol. 12.
15. McAuliffe M., Socolof M., Mihuc S., Wagner A., Sonderegger M. Montreal Forced Aligner: trainable text-speech alignment using Kaldi // Interspeech 2017: Conference Proceedings. 2017. <https://doi.org/10.21437/Interspeech.2017-1386>
16. Oppenheim A. V., Schaffer R., Yuen C. Digital Signal Processing // Systems, Man and Cybernetics. 1978. № 2.
17. Peterson G. H., Barney H. L. Control methods used in a study of the vowels // Journal of the Acoustical Society of America. 1952. Vol. 24 (2).
18. Stevens K. N. Acoustic phonetics. Cambridge: MIT Press, 1998.
19. Stone M. A guide to analysing tongue motion from ultrasound images // Clinical linguistics and phonetics. 2005. № 19 (6-7).
20. Taylor P. Text-to-Speech Synthesis. Cambridge: Cambridge University Press, 2009.
21. Tillmann H. G. Experimental and Instrumental Phonetics: History // Encyclopedia of Language and Linguistics / ed. by K. Brown. Amsterdam, 2006.

Информация об авторах | Author information

RU

Гончарова Оксана Владимировна¹, к. филол. н., доц.

¹ Российский университет дружбы народов им. П. Лумумбы, г. Москва

EN

Oxana Vladimirovna Goncharova¹, PhD

¹ Peoples' Friendship University of Russia named after Patrice Lumumba, Moscow

¹ Goncharova_oxvl@pfur.ru

Информация о статье | About this article

Дата поступления рукописи (received): 12.07.2025; опубликовано online (published online): 19.08.2025.

Ключевые слова (keywords): цифровая обработка речевого сигнала; автоматическое распознавание речи; инструментальные методы фонетики; глубокие нейронные сети; синтез и анализ речи; digital speech signal processing; automatic speech recognition; instrumental methods of phonetics; deep neural networks; speech synthesis and analysis.