

МАТЕРИАЛЫ И СООБЩЕНИЯ

БОНДАРКО Л. В., ЕГОРОВ С. К., ЛЮБЛИНСКАЯ В. В.,
НУРУМБЕТОВА Г. А.

О ПРИМЕНЕНИИ ВЫЧИСЛИТЕЛЬНОЙ ТЕХНИКИ В ЭКСПЕРИМЕНТАЛЬНО-ФОНЕТИЧЕСКИХ ИССЛЕДОВАНИЯХ

Использование ЭВМ в лингвистических исследованиях не только освобождает от необходимости вручную производить трудоемкую и однообразную работу, но и позволяет получать качественно новые данные, необходимые для развития науки о языке.

В последние годы вычислительная техника широко используется в экспериментально-фонетических исследованиях как за рубежом, так и в нашей стране. Однако с методикой такой работы, ее трудностями и принципиальными возможностями знаком довольно узкий круг специалистов. Это положение должно быть исправлено, оно требует регулярного обсуждения в печати, поскольку распространение вычислительной техники встречается не только с организационно-техническими, но и психологическими трудностями.

В настоящей работе описываются основные этапы одного из возможных методов спектрального анализа звуков с помощью ЭВМ. Цель работы — показать на конкретном материале как общение экспериментатора с ЭВМ, так и возможности, вытекающие из этого общения. Весь иллюстративный материал, приводимый в статье, — это спектрограммы звуков каракалпакского языка (частично — киргизского и узбекского языков); он получен в ходе исследования губной гармонии гласных в каракалпакском языке.

Обычно сведения об акустических характеристиках речевых сигналов в фонетических исследованиях получают в результате анализа спектров — как динамических, дающих представление об изменении частотных составляющих во времени, так и мгновенных, дающих сведения о среднем амплитудно-частотном спектре на некотором интервале длительности.

Современные традиционные установки для спектрального анализа состоят, как правило, из двух частей: аналогового спектрального анализатора (системы полосовых фильтров) и регистрирующего устройства, необходимого для получения динамической спектрограммы или спектральных срезов (амплитудных огибающих спектра в заданные моменты времени) на специальной бумаге или на фотопленке. Такое представление спектра предназначено для качественного анализа или для приближенного количественного, который позволяет сделать измерение линейных размеров изображения вручную.

В статье описывается способ получения спектральных характеристик речевых сигналов, где ЭВМ выступает как основное средство точного количественного анализа. Это переводит фонетические исследования на новый, более высокий уровень в сравнении со сложившимися традициями.

Схема установки, использовавшейся нами для спектрального анализа, приведена на рис. 1. Основные этапы и характеристики анализа заключа-

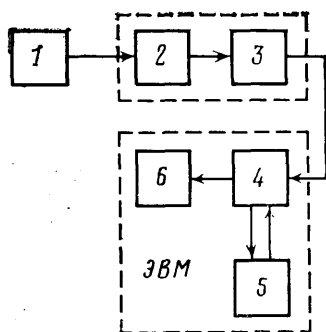


Рис. 1. Блок-схема установки для анализа акустических характеристик гласных (1 — магнитофон, 2 — спектральный анализатор, 3 — аналого-цифровой преобразователь, 4 — центральный процессор и оперативная память ЭВМ, 5 — лентопротяжный механизм, 6 — АЦПУ).

ются в следующем. Исходный речевой сигнал, предназначенный для анализа, записывается на высококачественный звуковой магнитофон (1). Полученные записи служат входными для внешнего по отношению к ЭВМ аналогового спектрального полосового анализатора (2). Анализатор содержит 46 параллельных каналов, в каждом из которых осуществляется фильтрация, выпрямление и сглаживание поступающего на вход сигнала¹. Средние частоты фильтров выбраны в диапазоне примерно от 110 до 10 000 Гц и имеют переменную ширину полосы от 60 до 500 Гц. Сигналы с выходов каналов анализатора с помощью коммутатора последовательно поступают на вход аналого-цифрового 12-ти разрядного преобразователя (АЦП) (3), подключенного к ЭВМ М4030-1 (4). Ввод сигналов в ЭВМ осуществляется при частоте квантования 10 кГц. Этот ввод в ЭВМ обеспечивает запись («перекачку») сигналов во внешнюю память машины, в качестве которой использовался накопитель на магнитной ленте (5). Результаты анализа выводились на алфавитно-цифровое печатающее устройство (АЦПУ) (6) в виде графиков и таблиц.

В описываемой системе анализа для обеспечения ввода и последующей обработки речевых сигналов был создан комплекс программ. Состав и содержание комплекса определялись доступными возможностями ЭВМ, бывшей в нашем распоряжении, ее конфигурацией. В этой статье мы в общем виде попытаемся описать возможности отдельных программ. (Программы написаны на языке PL/1 и FORTRAN-4.)

Первая программа, используя преобразователь, вводит речевые сигналы в оперативную память ЭВМ: преобразует их в последовательности целых чисел и в таком виде «перекачивает» во внешнюю память. Полученное представление речевых сигналов на новом носителе служит входным материалом для второй программы.

Вторая программа диагностирует качество записанных сигналов, выбирает среди них бездефектные, обозначает их и приводит к окончательному виду. В такой форме они записываются на новую магнитную ленту. Именно эта лента становится входным материалом для всех остальных программ комплекса. Диагностика и выбор — это выделение и обход плохих записей и возможная коррекция дефектных записей. (Дефектные записи обычно имеют место при вводе исходного материала низкого качества на звуковом магнитофоне или из-за помех в устройствах ввода.) При переносе на новую ленту остаются сигналы, удовлетворяющие выбранным критериям.

Есть смысл более подробно рассмотреть структуру записи данных, получаемых как результат второй программы. Каждому образу сигнала предшествует служебная запись, содержащая его атрибуты (инициаль). В инициаль входят следующие атрибуты, описывающие сигнал: название сигнала, имя диктора, место и дата осуществления записи, имя экспериментатора, параметры преобразователя и т. д. Создание инициали необходимо не только для текущих исследований (оформление результатов в удобной форме), но, главным образом, для организации архивов данных. При создании архива наличие инициали позволит осуществлять его упорядочивание и выбор по различным ключам, работу с каталогом.

В описываемой работе записи сигналов на магнитной ленте являются числовым представлением спектрограмм анализируемых звуков (фоном, слогов, слов, фраз, текстов). Самой распространенной задачей среди задач, встающих при обработке речевых сигналов, является визуальный анализ спектрограмм. Представление спектрограмм с помощью ЭВМ может быть реализовано многими способами и зависит от имеющейся в наличии техники вывода (АЦПУ, графопостроитель и т. д.).

Проще всего отпечатать массив чисел, составляющих спектрограмму, т. е. выходные напряжения каналов анализатора. Спектрограмма в памяти ЭВМ хранится в виде таблицы, число строк которой определяет число каналов спектрального анализатора, а число столбцов — число временных отсчетов (срезов). Но работа с такими таблицами технически неудобна и

¹ Аналого-цифровой преобразователь в комплекте с анализатором был разработан в ВЦ Вильнюсского государственного университета.

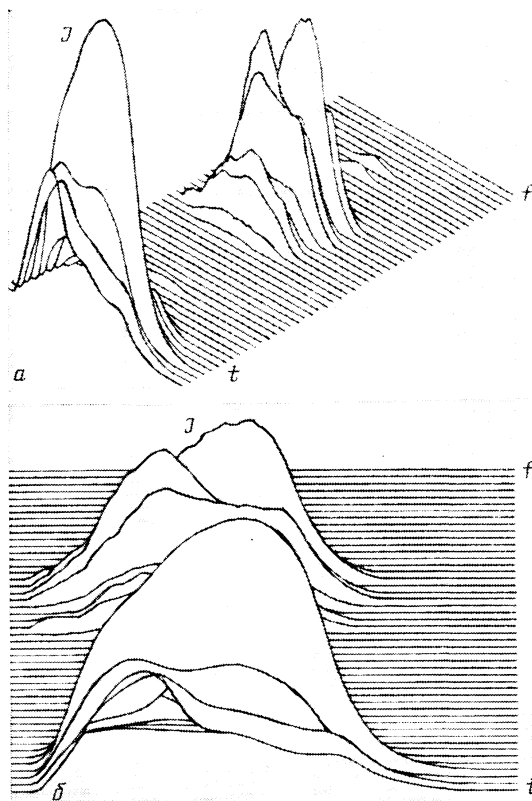


Рис. 2. Спектрограммы гласного [i], полученные на графопостроителе. Ось частот — f , ось времени — t , ось интенсивности — I .

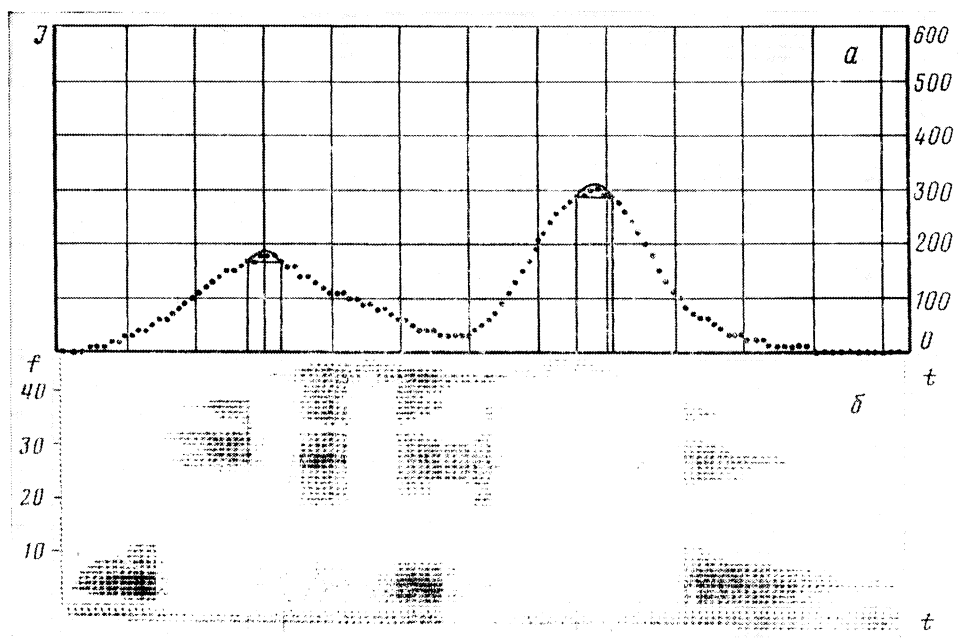


Рис. 3. Образец тоновой спектрограммы, получаемой программным способом: а) огибающая интенсивности слова; б) динамическая спектрограмма того же слова. На оси абсцисс — длительность (t), на оси ординат интенсивность (I) на рис. а) и частота (f) на рис. б).

малоеффективна. Графическое представление спектрограмм оказывается намного удобнее.

При наличии графопостроителя есть возможность получить аксонометрическое изображение динамической спектрограммы. В этом случае по одной оси откладываются значения частот фильтров анализатора, по второй оси откладывается значение времени, а по третьей оси — значения интенсивности (выходного напряжения каналов).

Рис. 2 (а и б) изображает спектрограммы звука [л], полученные на графопостроителе при двух различных поворотах осей.

При отсутствии графопостроителя для графического представления спектрограмм может быть использовано АЦПУ. Здесь, работая с литерами различной плотности (степени зачернения на поверхности листа бумаги), удается получить на двумерной картинке имитацию трехмерного изображения.

Рис. 3 дает пример спектрограммы, полученной с помощью АЦПУ для слова [ӨЗБЕК]. На картинке видно, что наибольшее затемнение соответствует участкам максимальной интенсивности, которое передается набором литер с наибольшей плотностью.

Так как изображение спектрограммы не использует целиком поверхности листа, на оставшейся части картинке (верхняя половина) печатается график изменений во времени усредненных по всем каналам значений выходных напряжений. Этот график позволяет составить представление об изменении интенсивности речевой волны и помогает анализировать спектрограммы: в частности, выделять стационарные и переходные участки гласного, выделять ударный гласный и т. д. Например, в приведенной на рис. 3 спектрограмме ударный и безударный гласный легко различимы по кривой интенсивности.

На картинке временная ось оцифрована, поэтому на ней легко фиксируются моменты, соответствующие границам сегментов, максимумам и другим особым точкам спектрограммы.

Возможность локализовать во времени особые точки и получить их числовое представление позволяет осуществить анализ спектральных срезов. Для этой цели предназначена третья программа разработанного комплекса. Она позволяет получить на одной картинке изображения от 1 до 5 «спектральных срезов» — зависимостей интенсивности выходных сигналов в частотных каналах от средней частоты фильтров — для одной записи. Графики строятся на 46 точек по оси абсцисс (центральные частоты фильтров) и на 60 точек по оси ординат (интенсивность в дБ от 10 до 70). Каждая кривая может представлять либо единичный спектр (срез за один временной отсчет), либо спектр, усредненный за несколько последовательных временных отсчетов (срез пучка).

На рис. 4 приведен образец спектральных срезов, полученных для слова [ӨЗБЕК]; здесь сплошной жирной линией соединены точки, соответствующие спектральному срезу на стационарном участке первого гласного, полученного путем усреднения данных за 5 отсчетов; прерывистая линия — средний спектральный срез стационарного участка второго гласного, усредненного за 7 отсчетов; сплошная тонкая линия — единичный срез в области переходного участка на границе согласного [b] и гласного [e].

Заметим, что различия в спектрах первого и второго гласного, которые можно обнаружить и на рис. 3, здесь могут быть оценены более определенно: огубленный [ö] характеризуется большей интенсивностью первого спектрального максимума и меньшей интенсивностью частотных составляющих в области F II, F III и более высоких частот. Эти факты могут быть интерпретированы как свидетельство существенных различий в качестве обоих гласных слова: отмеченные отличия могут являться результатом огубленности первого гласного и отсутствия огубленности при произнесении второго гласного. В то же время первый гласный может быть оценен как открытый и передний (по частотам F I и F II), а второй — как более закрытый (по значению F I) и несколько более передний (по значению F II). Переходный участок от согласного [b] к гласному [e] характе-

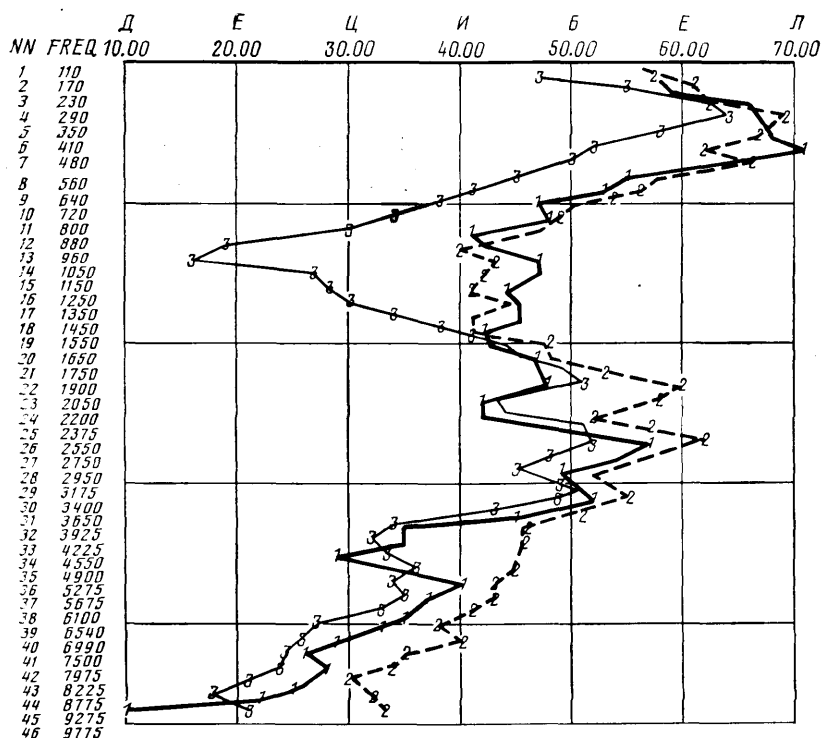


Рис. 4. Спектральные срезы стационарных и переходного участков гласного, динамическая спектрограмма которого приведена на рис. 3. На оси абсцисс — частота (F), на оси ординат — интенсивность (I). Жирная линия — стационарный участок первого гласного, тонкая линия — стационарный участок второго гласного, прерывистая линия — переходный участок от согласного к стационарному участку второго гласного.

ризуется низкой частотой F I (230 Гц) в начальной точке и высокой частотой F II (около 2750 Гц) в начале, что делает его очень похожим на i-образные переходы, возникающие на границе гласных с предшествующими мягкими согласными. Таким образом, сравнение огубленного гласного [ö] со вторым гласным того же слова по их спектральным срезам позволяет выявить существенные различия между этими гласными.

В ходе исследования каракалпакских гласных необходимо было выяснить степень влияния первого гласного (огубленного или неогубленного) на второй гласный в слове, поэтому потребовалось сравнение спектральных срезов гласных, находящихся в разных словах. Это сравнение оказалось возможным сделать с помощью четвертой программы комплекса, которая выбирает из разных записей спектральные срезы в указанные моменты времени (единичные срезы или средние в «пучке») и изображает их в виде соответствующих кривых на одном графике аналогично тому, как это делается в предыдущей программе. В этой программе сравнивались непервые гласные в словах с огубленным и неогубленным первым гласным. На рис. 5 приведены спектральные срезы вторых гласных из слов [buryl] и [baryl]. На рис. 6 приведены спектральные срезы вторых гласных в словах [wögiz] и [ʒegiz]. Сравнение огибающих этих спектров показывает, что гласный [y] из первого слова может быть охарактеризован как огубленный по сравнению с гласным [y] из второго слова: низкочастотный максимум имеет большую интенсивность и более низкую частоту, а частотные составляющие выше 1000 Гц, наоборот, характеризуются меньшей интенсивностью. В области F III имеются различия, свидетельствующие о том, что [y] из слова [buryl], имеющий более низкую частоту F III, является огубленным.

Для гласных [i] из второй пары слов соотношение несколько иное: [i] из первого слова может быть охарактеризован как более открытый (по частоте F I) по сравнению с [i] из второго слова. Наличие максимума

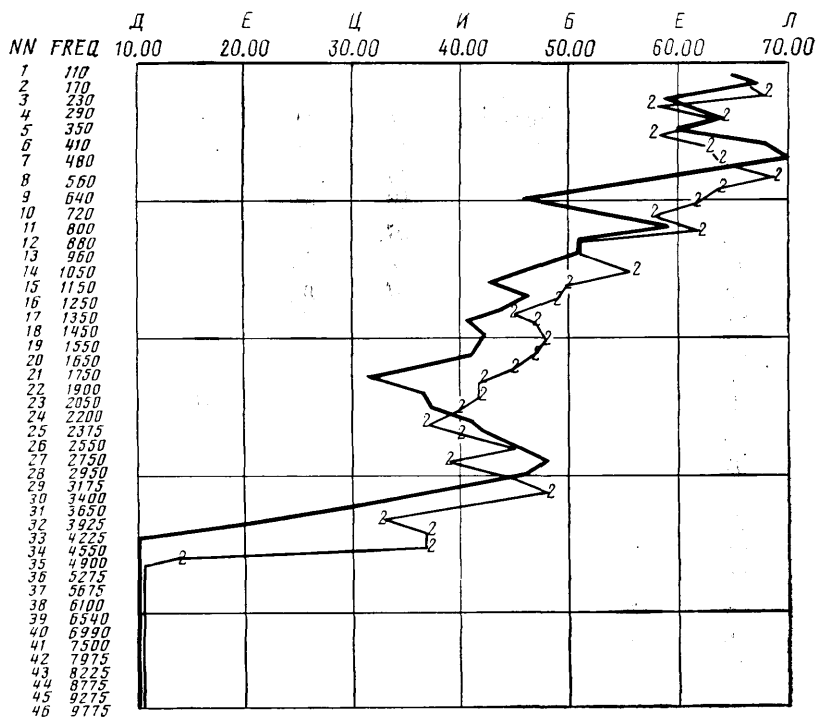


Рис. 5. Спектральные срезы гласных [y] из слов [buryl] и [baryl]. Сплошной тонкой линией — огибающая для неогубленного, сплошной жирной линией — огибающая для огубленного.

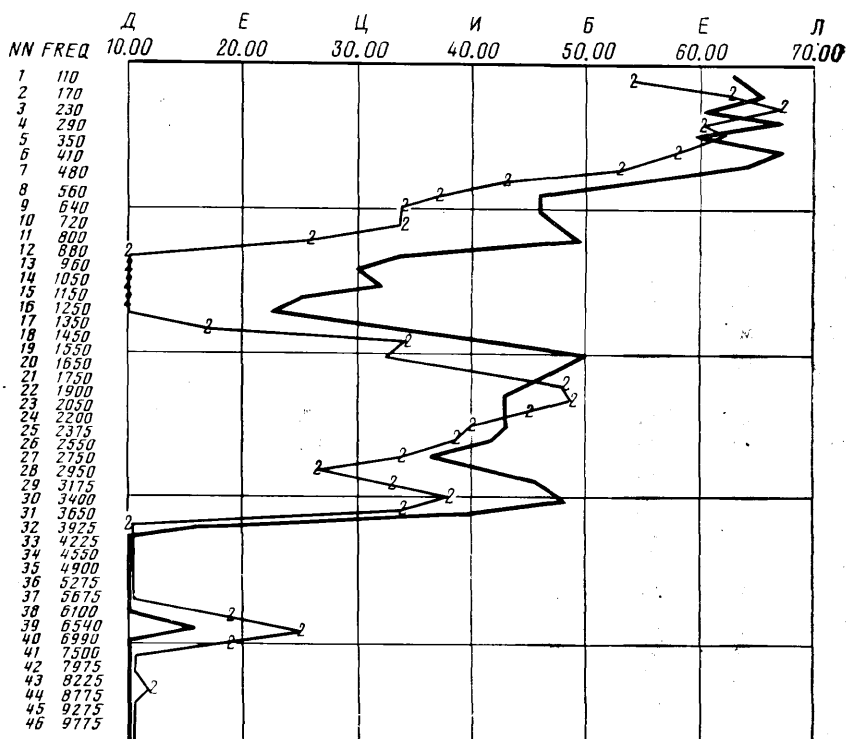


Рис. 6. Спектральные срезы вторых гласных из слов Θ ГПЗ [ögiz] — (сплошная жирная линия) и ЕГПЗ [jegiz] — (сплошная тонкая линия).

в области 640 Гц свидетельствует о возможной лабиализованности этого гласного. Об этом говорит и более низкое расположение F II по сравнению с [i] из слова [jegiz] (1650 Гц и 2050 Гц соответственно).

Для описываемой здесь системы анализа существенно, что программы позволяют получить на одной картинке не только срезы гласных, находящихся в одном слове, но и аналогичные характеристики гласных, находящихся в разных словах. Рассмотрим эти возможности на примере. Известно, что каракалпакский язык занимает промежуточное положение по степени выраженности губной гармонии между языками с ярко выраженной гармонией и с отсутствием таковой. В описываемом здесь исследовании в качестве языка с сильной губной гармонией был взят киргизский язык, а в качестве языка с отсутствующей губной гармонией — узбекский. Для оценки степени проявления губной гармонии в каракалпакском производилось сравнение непервых гласных каракалпакских слов с непервыми гласными из аналогичных слов киргизского и узбекского языков.

На рис. 7 (а и б) приведены спектральные срезы гласных в киргизском слове $\Theta\Gamma\ddot{y}3$ [ögüz] (а) и в узбекском слове ХУКИЗ [hvkiz] (б). По характеристикам киргизского гласного можно судить о значительной его огубленности: низкий уровень составляющих в области F_{II} , наличие максимума в области около 1000 Гц, общая сравнительно низкая интенсивность по сравнению с гласным (огубленным же!) первого слога — все эти характеристики соответствуют общefonетическим представлениям об акустических коррелятах огубленности. На рис. 7 (б) второй гласный слова имеет совсем другие характеристики: это типичный закрытый гласный переднего ряда, интенсивность высокочастотных составляющих очень большая, так что здесь нельзя говорить о какой-либо лабиализованности. Хотя это с точки зрения традиционных представлений о свойствах тюркского вокализма и тривиально, для общего хода исследования каракалпакского языка возможность сравнивать спектральные срезы гласных из разных языков очень существенна: сравнение рис. 6 и 7 (а и б) показывает, что каракалпакский гласный в слове [wogiz] по своим характеристикам действительно занимает промежуточное положение между киргизским лабиализованным и узбекским нелабиализованным гласным. Это, как видно из рисунков, касается относительного уровня интенсивности частотных составляющих в области 1650—3175 Гц. Можно предположить, что статистический анализ большого материала позволил бы количественно оценить степень проявления губной гармонии. Естественно, что для этого необходимо не только сравнение единичных спектров, которые мы описывали до сих пор, но и получение соответствующих статистик.

Для статистической обработки материала была создана программа, которая позволила получить усредненные спектры изучаемых гласных из разных слов. Эта программа выбирает указанные срезы из разных записей (номера записей и срезов в них указывает исследователь на основании предварительного анализа спектрограмм) и вычисляет среднее значение интенсивности по каждому каналу, таким образом получая усредненный спектр. Одновременно определяются значения минимального и максимального элементов для каждого канала. Полученные значения выводятся в виде таблицы и графика в той же шкале, что была сделана для двух предыдущих программ.

Графики служат для наглядного представления результатов, а вычисленные значения могут быть использованы в качестве исходных для дальнейшей элементарной статистической обработки (например, выявление различий по заданным критериям и т.д.). На рис. 8 (а и б) приведены в качестве образца средние спектры каракалпакских гласных /а/ и /о/. Существенно, что рядом со спектрограммой может быть напечатан список слов, содержащих данный гласный, а также числовые значения интенсивности в каждом из каналов.

Поскольку целью данной работы является в основном описание методики машинной обработки речевых сигналов, мы не останавливаемся подробно на акустических характеристиках огубленности в каракалпакском языке: это — тема для специального детального описания. Здесь мы хотели бы охарактеризовать как этапы уже проведенной работы, так и дальнейшие перспективы исследований подобного рода.

Выше мы перечислили содержание программного комплекса примени-

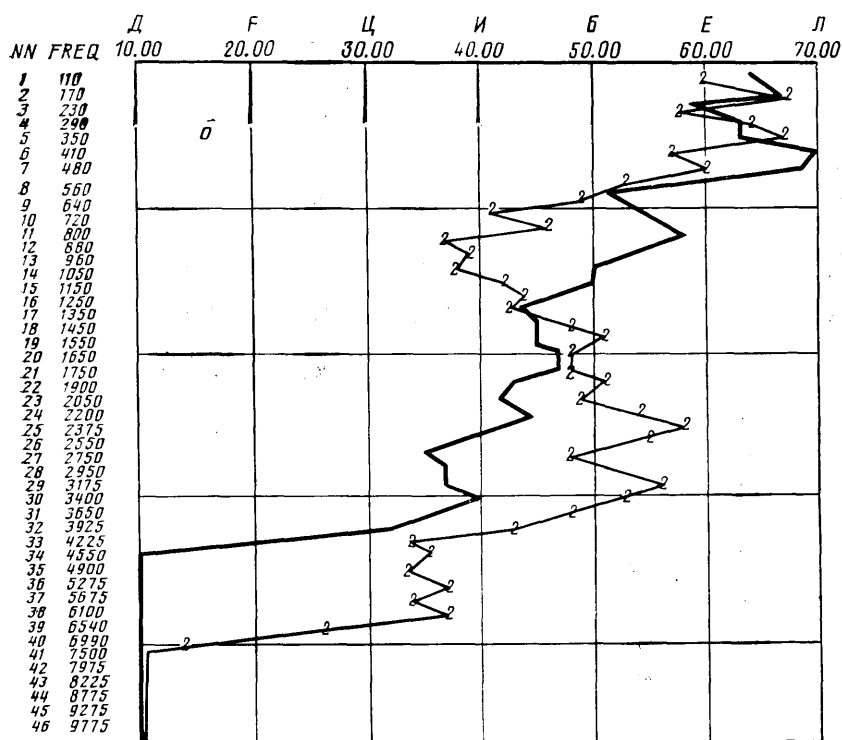
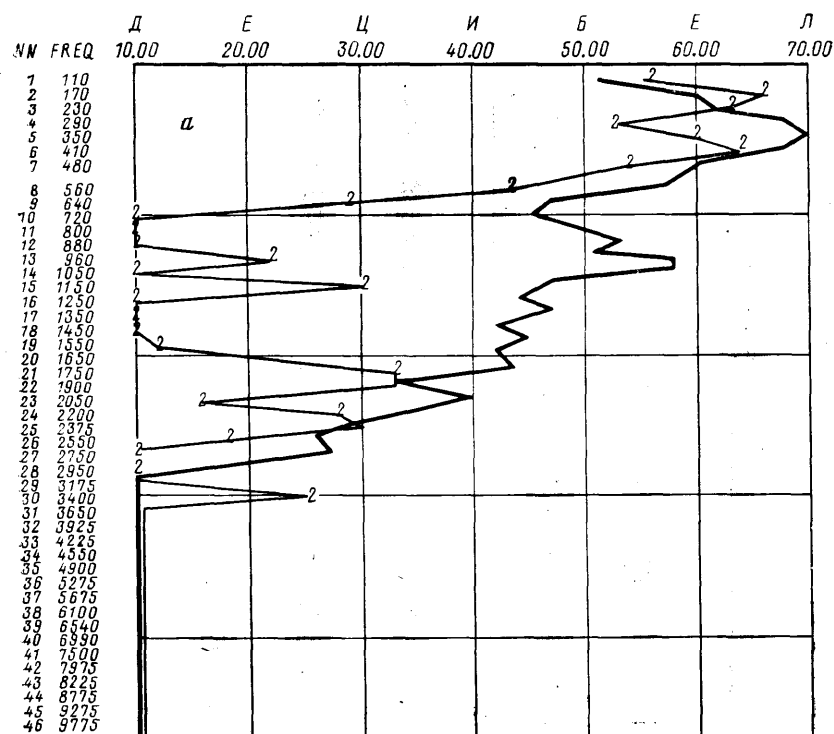


Рис. 7. Спектральные срезы гласных из слов — а) киргизский **ӨГҮЗ** [ögüz] (жирная линия — гласный [ö], тонкая линия — гласный [ü]), б) узбекский **ХҲКИЗ** [hökiz] (жирная линия — гласный [ö], тонкая линия — гласный [u]).

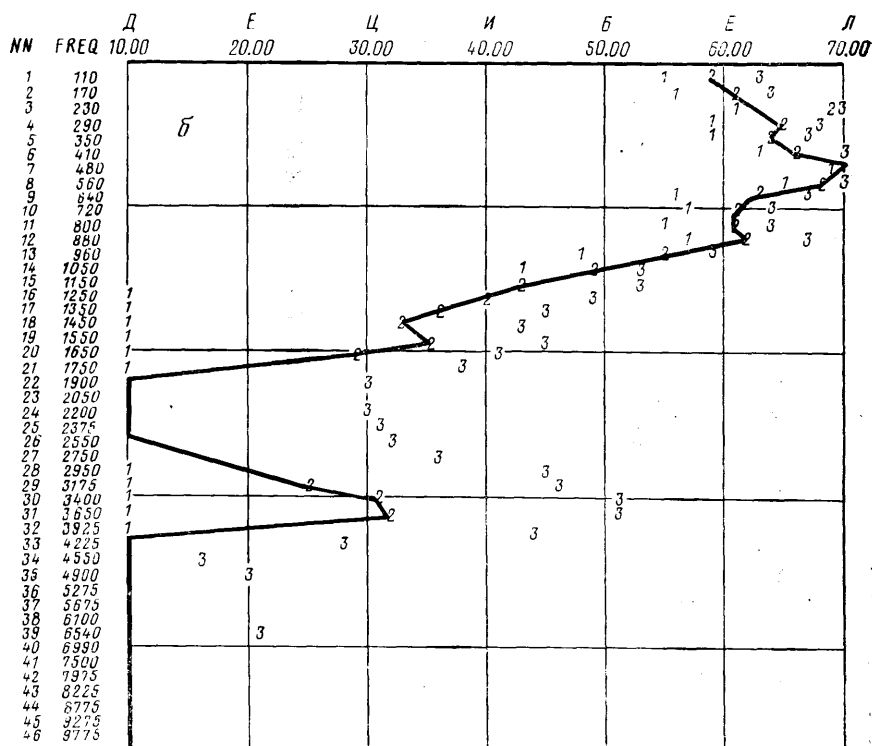
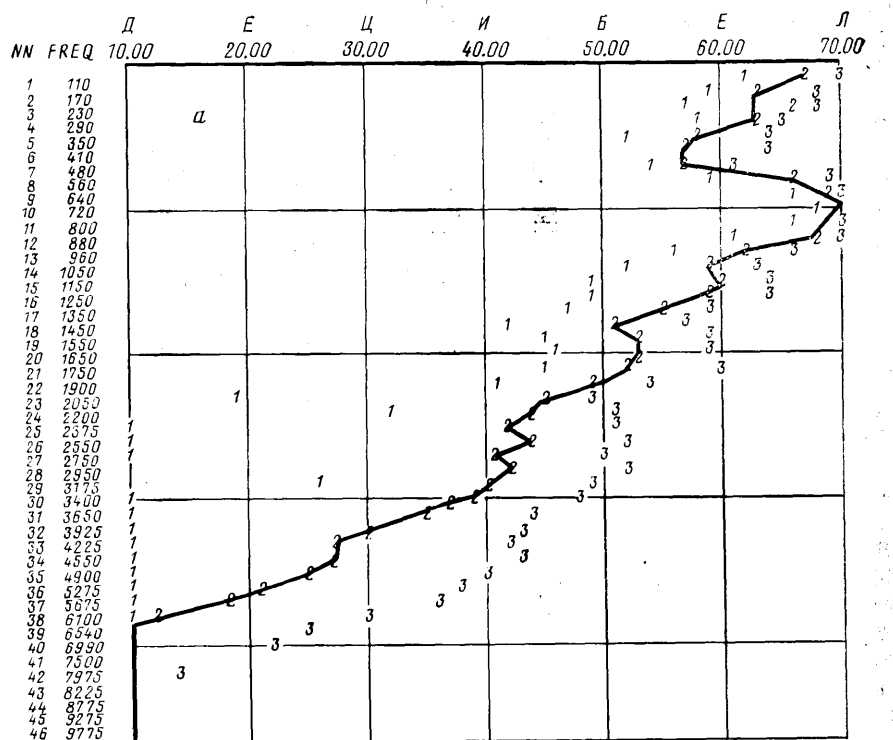


Рис. 8. Средние спектры гласных а) /а/ и б) /о/. Цифрой 1 обозначены минимальное значение, цифрой 2 — среднее, цифрой 3 — максимальное.

тельно к одному из конкретных примеров фонетического исследования речевых сигналов. Следует сказать, что среди данного комплекса первые три программы являются базовыми и используются для всех задач, связанных с обработкой речевого спектра. Остальные программы ориенти-

рованы на одну задачу, разработаны специально для нее, но могут быть использованы для аналогичных целей. Все программы параметризованы, написаны по возможности в общем виде и управляются параметрами, задаваемыми применительно к поставленным исследователем задачам.

Попробуем коротко сформулировать в общем виде, насколько разумно использование вычислительной техники в фонетических исследованиях и в чем заключаются преимущества такого подхода, что он дает нового по сравнению с традиционными методиками, что предлагает машина фонетисту вместо привычной работы с маловыразительной картинкой.

Основной машинной обработки является возможность численного представления речевых сигналов, которая с помощью современных технических устройств (аналого-цифровых преобразователей) осуществляется с достаточно высокой точностью. Это в свою очередь допускает применение более тонких методов численной обработки речевых сигналов.

Размеры памяти современных ЭВМ позволяют хранить огромные массивы информации и организовать архивы данных, с другой стороны, высокое быстродействие ЭВМ позволяет обрабатывать информацию в количествах, ранее считавшихся недоступными. Преимущества машинных методов анализа в сравнении с почти качественными методами оценки параметров речевых сигналов с помощью линейки и карандаша очевидны.

Несмотря на большие преимущества машинных методов, в фонетике они пока еще не находят широкого применения. Для более интенсивного внедрения этих методов необходимо решить ряд проблем.

На первое место, несомненно, должно ставиться требование обеспечения техническим оснащением коллективов, работающих в области фонетики. Нужны машины (ЭВМ) с необходимым комплектом внешних устройств, включающие современные средства диалога и вывода данных (графопостроители, графические и полутоновые дисплеи) и т. д.²

Второй фактор, но не менее важный — создание коллективов специалистов, работающих в этой области. Здесь могут возникнуть трудности не только организационного, но и психологического характера. Дело в том, что выполнение работ, связанных с обработкой речевых сигналов на ЭВМ, практически недоступно одному специалисту, а требует взаимодействия нескольких. Так, например, сделанная нами работа объединила усилия лингвистов, программиста, математика и двух инженеров (специалиста по вычислительной технике и специалиста по исследованию речи). Работа выполнена в результате сотрудничества кафедры фонетики Ленинградского университета с Лабораторией физиологии речи и с Вычислительным центром Института физиологии им. И. П. Павлова АН СССР.

В предлагаемой ниже таблице мы попытались представить в систематизированной форме основные этапы проведенной работы, средства, которыми она осуществлялась, и состав участников.

Как видно из таблицы, место экспериментатора-фонетиста определяется не только его ролью при выборе исходного материала или при анализе и интерпретации данных. Он принимает участие при создании алгоритма обрабатывающей программы, при составлении задания для нее.

Таким образом, использование ЭВМ в фонетических исследованиях возможно только при условии постоянного контакта фонетиста-экспериментатора со специалистами в области вычислительной техники. Опыт разработки описанной в статье методики показывает, что для инженера или программиста нет принципиальных трудностей, связанных со спецификой представленного материала. Наоборот, экспериментатор-фонетист в ходе работы сталкивается с новыми проблемами и понятиями. Поскольку именно фонетист ставит исследовательскую задачу и определяет конкретные этапы ее решения в содержательном смысле, его способность понять требования и возможности машинного эксперимента приобретает решающее значение. Это означает, что вопрос о подготовке лингвистов — спе-

² Видимо, наиболее реальный путь для лингвистов, не имеющих собственного технического обеспечения, это работа на базе ВЦ других научных или учебных подразделений.

Вид работы	Используемые ресурсы	Участники
1. Постановка задачи и выбор средств для ее решения с учетом доступных средств	Профессиональные возможности участников, конфигурация ЭВМ и средства, обеспечивающие звукозапись, состав базового программного обеспечения	Лингвисты, инженеры, математики, программисты
2. Подготовка речевого материала для записи и анализа	Звуковой магнитофон высшего класса по качеству	Фонетист-исследователь, дикторы, специалист по звукозаписи
3. Запись сигналов во внешнюю память ЭВМ	Звуковой магнитофон, аналого-цифровой преобразователь, ЭВМ, базовое программное обеспечение	Инженер, программист
4. Диагностирование качества полученных записей, перезапись с включением инициализации и контрольное прослушивание сформированных записей	ЭВМ, цифро-аналоговый преобразователь, базовое программное обеспечение	Программист, инженер
5. Обработка записей речевого сигнала на ЭВМ с целью предварительных данных для визуального анализа (динамические спектрограммы и т.д.)	ЭВМ, графический дисплей, графопостроитель и т.д. (средства графического отображения результатов), базовое программное обеспечение	Фонетист-исследователь, дикторы, математик, программист
6. Предварительный анализ результатов. Переход к обработке на следующем уровне (уточнение параметров, элементарная статистика)	Данные средств графического отображения результатов (листинги с АЦПУ, картинки с графопостроителя и изображения на экране дисплея)	Фонетист-исследователь, математик
7. Обработка с использованием программ более высокого уровня. Общая интерпретация полученных результатов	ЭВМ, специально написанные программы	Фонетист-исследователь, математик, программист

циалистов такого профиля очень важен. Назрела необходимость обеспечить такую подготовку путем разработки специальных программ, создания учебных пособий и руководств. Интерес к фонетическим исследованиям в нашем многонациональном государстве может быть поддержан теми средствами, которые дает исследователю вычислительная техника.

С использованием этих средств не только укрепляется и становится более совершенной экспериментальная база уже имеющихся в стране фонетических лабораторий, но и облегчается задача создания новых: сегодня едва ли целесообразно приобретать аналоговую измерительную аппаратуру, тем более, что в серийном производстве многих необходимых приборов нет.

В этой статье описан лишь один из возможных путей использования ЭВМ в экспериментально-фонетических исследованиях. Однако кроме исследования спектральных характеристик звуков ведутся и исследования просодических характеристик, и работы по сегментации и пересадке звуков. Уже разработаны программы для этого и можно надеяться, что в скором времени они станут известны широкому кругу исследователей.