

В. В. Мелконян, В. Н. Айрапетян

К ВОПРОСУ О СПЕКТРАЛЬНОМ АНАЛИЗЕ СОГЛАСНЫХ ЗВУКОВ АРМЯНСКОГО ЯЗЫКА

Вопрос о полном и всестороннем экспериментально-фонетическом анализе всего состава фонем армянского языка, в особенности его консонантной системы, стоит на повестке дня советской арменистики. Достижения современной науки и техники подвели научные изыскания в области общественных наук, в том числе и лингвистики, к новой ступени исследовательских работ, в которой наблюдается всемерное использование методов количественного анализа, особенно в работах экспериментального характера.

В данной статье описывается метод спектрального анализа, который разработан в кабинете экспериментальной фонетики ИЯ АН Арм. ССР. На основании некоторых примеров излагается суть предложенного метода.

В области экспериментальной фонетики применяется несколько методов объективного анализа звуков речи. Одним из них, наиболее совершенным, является спектральный анализ. Современная экспериментальная фонетика при помощи данного метода старается дать наиболее полное описание звуков.

Из теории информации известно, что объект тем полнее описан, чем большим количеством информации о нем мы располагаем.

В предложенном нами методе звук описывается тремя координатами: частота — время — интенсивность.

В настоящий момент перед кибернетикой стоит задача «распознавания образа». В число этих образов входит и звуковой образ — фонема.

Таким образом, данные спектрального анализа звуков речи кроме сугубо лингвистических целей используются и в кибернетике (автоматы, управляемые речью, «автоматический секретарь», устный машинный перевод и др.).

По анализу гласных имеется достаточное количество исследований. Но при анализе согласных экспериментаторы сталкиваются с трудностями. Спектр согласных звуков более сложен в связи со спецификой данных звуков, он приближается к спектру шума; труднее определить закономерности для каждого отдельного звука. Согласный звук невозможно исследовать изолированно. С этой целью производится анализ звуко сочетаний, где участвует данный звук. Как преодолеваются эти трудности в предлагаемом методе, видно из примеров, разбираемых в данной статье.

Несколько слов о методике проведения опыта. Звукосочетание или отдельное слово записывалось на непрерывное кольцо, которое проходило мимо головок магнитофона, т. е. мы имели циклически повторяющееся звукосочетание (в опытах использовался магнитофон «МЭЗ-15»). Далее запись снималась с выхода усилителя воспроизведения и подавалась на поочередно включаемые входы фильтров анализатора. Выходы фильтров были запараллелены и подключены к осциллографу, снабженному фотоприставкой. Вертикальная ось осциллографа была калибрована в децибеллах.

Одним из основных требований к нашей разработке являлось обеспечение точной синхронизации анализируемого материала по временной оси, ибо любое смещение во времени могло привести к искаженным результатам.

Синхронизация осуществляется следующим способом: сетка тиратрона генератора развертки через сопротивление 1,2 мом соединяется с контактом (рис. 1), укрепленным на верхней панели магнитофона. Другой контакт (2) соединен с выходом «земля» электронно-лучевого осциллографа. На кольцо магнитофонной ленты прикрепляется металлическая фольга так, что при подходе начала звукосочетания к головке воспроизведения фольга замыкает контакты, т. е. сетка тиратрона заземляется. Генерация срывается, и после прохождения фольги (мимо контактов) начинается развертка на экране осциллографа, причем с одной и той же точки, чем и обеспечивается синхронизация покадровой съемки анализируемого материала. Нами использовались фильтры от серийного спектрометра типа «СЗЧ».



Рис. 1. Схема синхронизации.

Время развертки при анализе устанавливается на большую длительность, чем продолжительность звучания исследуемого звукосочетания. В конце каждого ряда фотосъемок анализируемого материала на экран осциллографа подается от звукового генератора постоянная частота в 40 гц для установления масштаба временной характеристики каждого звукосочетания (слова или слога, в зависимости от характера исследования или цели его).

Основным диапазоном для спектрального анализа звуков речи (включая все группы согласных) можно, по-видимому, считать полосу частот до 20000 гц. Фильтры от анализатора «СЗЧ» охватывают диапазон частот до 20 кгц и даже более. Весь диапазон частот спектрометра разбит на 27 фильтров, в которых полосы пропускания частот установлены по логарифмической шкале. Полосы пропускания у высокочастотных фильтров довольно широки, что создает определенные неудобства при анализе шумных согласных. Однако и при таких условиях, путем многочисленных съемок, нам удалось довольно точно установить параметры для каждой группы согласных звуков речи (смычных, фрикативных и аффрикат, а также соноров). В дальнейшем предполагается разбить диапазон в 20 кгц на несколько фильтров с перестраиваемой полосой частот. Это позволит еще более повысить точность анализа.

В процессе анализа фильтры подключались поочередно. Звукосочетания, входящие в состав слова, сравнивались при разных частотах по временной оси. В результате анализа мы имели целый ряд кадров, отражающих спектральную картину целого сочетания звуков (слова).

При испытании метода спектрального анализа речевых сигналов первоначально подверглись изучению согласные звуки: ряд аффрикат и глухих придыхательных в окружении различных гласных. Потом последовательности типа *усу* (гласный+согласный+гласный) были заменены односложными, двусложными, а в ряде случаев и трехсложными словами. Было исследовано большое количество слов, отражающих все возможные позиции согласных звуков речи, входящих в состав разных групп (по способу и месту образования).

Для иллюстрации предложенного метода спектроанализа звуков речи, его релевантных смыслоразличительных единиц (фонем), отобрано два звукосочетания (слова) в произнесении двух дикторов — мужчины и женщины: *шшрпг* (*hajoγ*) и *шшшш* (*peχas*).

При сопоставлении отдельных кадров исследуемого звука составляется общее представление о нем. В ходе эксперимента было установлено, что общая полоса пропускания частот для мужского голоса охватывает диапазоны, начиная с полосы пропускания 4 фильтра анализатора, с параметрами 89,8 ÷ 133,4 гц до 24 фильтра, с полосой 9130 ÷ 11500 гц, а для женского голоса — преимущественно с 6 фильтра (142,5 ÷ 179,5 гц) до 24. При этом наблюдалось, что в полосе пропускания высокочастотных фильтров (22—24) отмечаются характерные выбросы спектральных линий на месте шумных звуков речи: глухих фрикативных и аффрикат. Согласные звуки в отобранных нами словах находятся либо в окружении гласных, либо предшествуют им или же находятся в позиции абсолютного конца слова после гласных (рис. 2—5). Окружающие гласные, где тоны превалируют над шумами, дают более яркую картину выбросов, по сравнению с согласными. Более того, известно, что гласные звуки речи имеют гармоники, которые в кратных выражениях периодически повторяются.

Кстати, отметим, что форманты гласных в наших опытах соответ-

ствовали данным, приведенным в предшествующих работах по их исследованию (работы Г. Оганесяна, А. Арамян и А. Хачатрян)¹.

Для гласных наблюдалась картина непрерывающихся выбросов частот от низкочастотных фильтров к диапазону средних частот, образующих целую область с довольно широким охватом частот. Таким образом, образуется довольно широкая область частот, обозначаемая некоторыми авторами² символом «В». По данным ряда фильтров составляются форматные картины (F-картины).

Что же касается согласных звуков, то их гармоники имеют аperiодический характер, трудно поддающийся анализу. Следует отметить, что не все согласные имеют в общей картине анализа аналогичные характерные выбросы. В диапазоне низких частот отмечается лишь основной тон гласных без характерных выбросов амплитуды. На месте согласных звуков пики не наблюдаются. В полосе последующих фильтров на месте согласного звука γ [j] отмечаются небольшие линии без намека на форманту. Область форманты для согласного γ [j] обнаруживается в диапазоне частот $7240 \div 9130$ гц по восстанавливаемой вертикали. Проведение вертикали по всем кадрам для определения приближенных границ отдельных звуков возможно в силу уверенности, что все кадры синхронизированны по времени. Это обстоятельство позволяет экспериментатору обозревать весь ряд кадров как единое целое.

Так, например, если по общей картине анализа слова հաջող [haʝoɟ] (муж.) наиболее высокий пик, характеризующий F-картину для гласного ա (а) наблюдаем в полосе 16 фильтра ($1433 \div 1812$ гц), то для гласной ո (о) область максимальной концентрации энергии отмечается в диапазоне $359 \div 453$ гц. Благодаря распределению частот фильтров по логарифмической шкале мы имеем довольно узкополосные фильтры в диапазоне низких, частично-средних частот. Следовательно, мы вправе говорить об областях концентрации энергии, именуемой формантой.

Необходимо отметить, что соотношение амплитудных характеристик, по которым отмечаются области формант для каждого звука речи, свойственно не только гласным. По общей спектральной картине мы определяли форманты и для согласных звуков (см. выше). Известно, что амплитуда при анализе звуков речи выражает не абсолютные, а относительные, в данном случае для каждого ряда, величины. По нашим опытам можно одновременно с определением частотных характеристик

¹ Г. А. Оганесян, Анализ гласных электроакустическим методом в зависимости от пола и регистра, применительно к армянскому языку, Ереван, 1950 (диссертационная работа); А. М. Арамян, Экспериментально-фонетическое исследование и классификация гласных фонем армянского литературного языка, Ереван, 1958 (диссертационная работа); А. А. Хачатрян, Спектральный и рентгенологический анализ гласных армянского и английского языков, Ереван, 1963, (диссертационная работа).

² Г. Фант, Акустическая теория речеобразования (перевод с англ.), издательство «Наука», М., 1964, стр. 14.

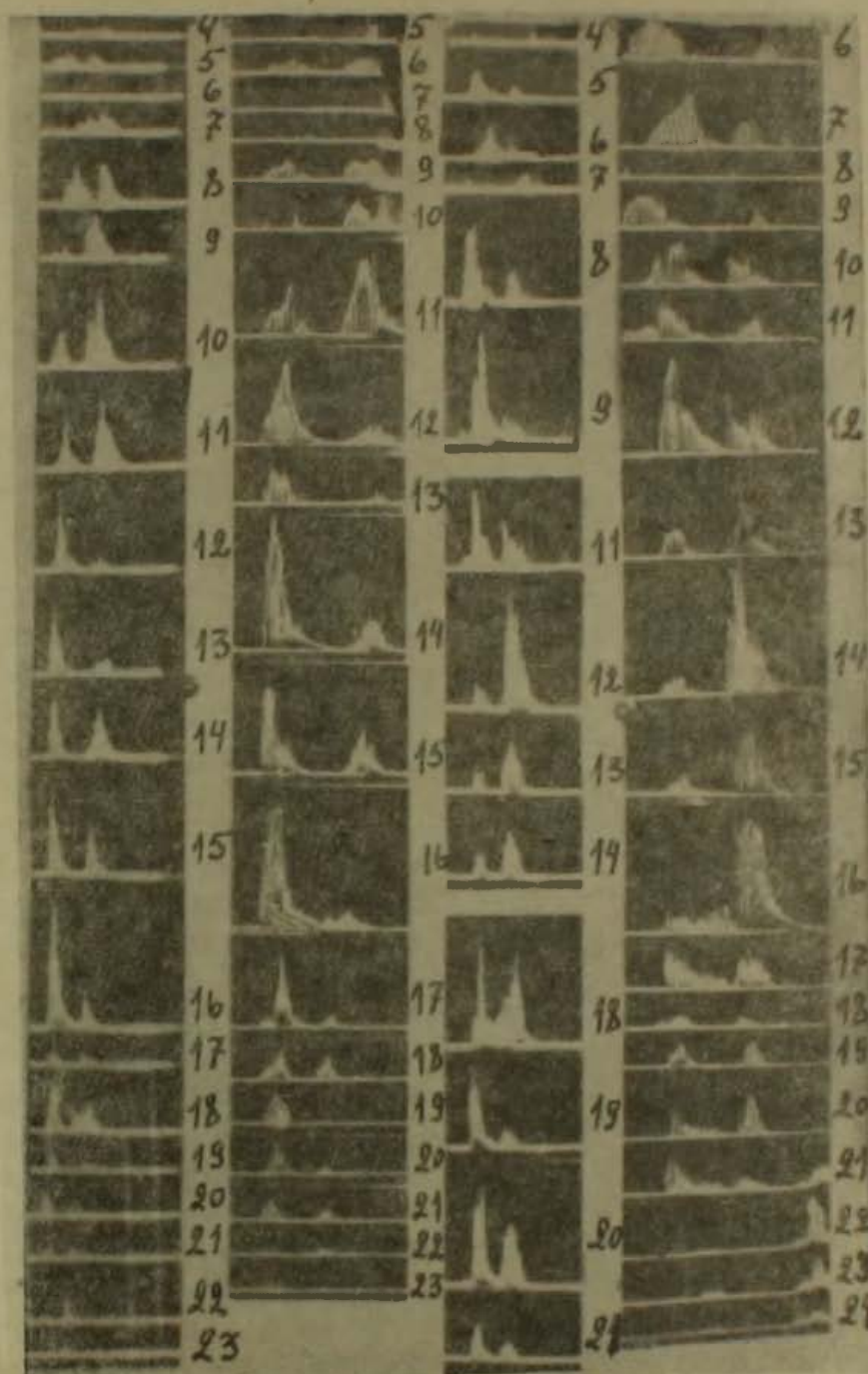


Рис. 2
слово
հաճօղ
(hadžo)
муж.

Рис. 3
слово
հաճօղ
(hadžo)
жен.

Рис. 4
слово
նեխս
(neχas)
муж.

Рис. 5
слово
նեխս
(neχas)
жен.

для каждого звука точно установить интенсивность по амплитуде.

При проведении эксперимента опрос фильтров проводился в прямой последовательности, а анализ производился после того, как по вертикальной оси определялись границы звуков. При определении этих границ учитывались характерные особенности звуков по способу образования. Данные спектрального анализа восполнялись данными лингвистического анализа, в результате чего получалось полное представление о физических характеристиках звуков речи. Например общеизвестно, что ряд согласных звуков характеризуется смычным элементом в начальной стадии произведения данного звука, далее следует эксплозия (взрыв), в результате чего наблюдаем реализацию указанных фонем, преимущественно смычно-эксплозивных, в последующем звуке. Акустическая характеристика взрывных, по определению некоторых авторов³, отмечается тем, что взрывные звуки речи имеют несколько характерных участков: нулевой — смычка, первый — взрывная часть и т. д. Согласно данной характеристике, можно отметить, что в речевой цепи при переходе от предшествующего звука к взрывным звукам речи и к смычно-эксплозивным (аффрикатам) отмечается спад спектральных линий (в случаях, если не наблюдается колебание голосовых связок). В первом случае, в слове հաջող [haʃoɣ] разрыва спектральных выбросов не наблюдалось, ибо согласные ղ [j] и ղ [ɣ] являлись звонкими, и колебательное движение связок не прекращалось.

Тем не менее, слияние первых гармоник наблюдалось лишь в диапазоне низкочастотных фильтров (4—7 в мужском произношении и 6—8 в произношении диктора-женщины).

По общей картине слова հաջող [haʃoɣ] (жен.) можно отметить следующие особенности: спектральная картина на низких частотах примерно повторяется, но со сдвигом «вверх» (рис. 2—3). В дальнейшем характерные пики согласных звуков, незначительные по амплитуде, отмечаются преимущественно в полосе высокочастотных фильтров. Здесь же отмечается относительная самостоятельность конечной согласной ղ [ɣ], которая характеризуется в диапазоне частот 906 ÷ 1141 гц и частично в полосе 1141 ÷ 1436 гц. Спектральные линии хоть и сливаются на границе с предшествующей гласной, но мы имеем некоторое изменение характера огибающей. По данным последующих фильтров, по результатам других анализов можно восстановить истинную картину частотных характеристик каждого звука речи. Мы особенно подчеркиваем это обстоятельство, анализ звуков мы можем считать тем полнее, чем больше будет подвергаться анализу данный звук речи.

В силу той особенности, что женский голос по тембру несколько выше мужского, по спектральной картине второго ряда слова հաջող [haʃoɣ] отмечаем некоторую сдвинутость «вверх» частотных спектров. Так, например, если характерные выбросы отмечаются в одном случае

³ М. А. Сапожков, Речевой сигнал в кибернетике и связи, Связьиздат, 1963, стр. 77.

в определенной полосе частот, то картина огибающей в произношении диктора-женщины повторяется в полосе частот, наблюдаемых выше.

Известно, что в задачи акустиков и фонетистов, занимающихся теорией речеобразования, входит точное определение резонансных частот звуков речи и выявление формантных полос. Именно поэтому анализируется речь в произношении многих дикторов, чтобы выявить не только параметры звуков речи, но и характер огибающей спектральных линий, чтобы по возможности исключить погрешности тембральной окраски.

Визуальный анализ двух последующих рядов спектральных данных слова *həʃuʃd* [пешас], позволяет нам четче разграничить параметры звуков речи в данном сочетании, поскольку здесь имеются глухие звуки. Например, по данным первых фильтров отмечаются едва заметные выбросы на месте гласных *h* [e] и *u* [a], а в полосе 5 фильтра 113,4÷142,5 гц. По мере увеличения интенсивности гласной *h* [e] отмечается незначительная по интенсивности полоса спектральных линий на месте фрикативного звука *ʃ* [x]. На последующих фильтрах, вплоть до 453 гц, интенсивность данного звука попеременно то увеличивается, то уменьшается, не давая, однако, характерного пика частот. Следует отметить, что большая часть характерных выбросов для согласных звуков ориентируется по характерным пикам окружающих гласных звуков. Если по общей картине спектра отметить довольно высокие пики, характеризующие гласные звуки, то отчетливо будут выделяться форманты как для первой гласной, так и для второй. На частотах 7240÷11500 гц выделяются преимущественно согласные звуки *ʃ* [x] и *ʒ* [c]. При этом можно вычленить для конечной аффрикаты *ʒ* [c] формантную полосу в диапазоне 9130÷11500 гц, которая могла быть тем точнее, чем уже были полосы пропускания у высокочастотных фильтров.

Фрикативный звук *ʃ* [x] имеет менее характерную спектральную картину, чем аффриката *ʒ* [c]. Но фрикативный звук отмечается довольно четко в двух поясах частот, в первом случае — в диапазоне 113,4÷179,5, после чего превалируют частотные пики гласной *h* [e]. Второй пояс для фрикативного звука начинается в полосе частот от 3620 гц и продолжается вплоть до частот 7240÷9130 гц (полоса пропускания 23 фильтра). Что же касается аффрикаты *ʒ* [c], то она имеет характерные выбросы в поясе частот от 4560 до 11500 гц, имея при этом выбросы формантного характера в диапазоне 9130÷11500 гц. Границы звуков речи отмечались нами по отмеченному выше принципу. Смычно-фрикативная *ʒ* [c], благодаря элементу смычки, четко разграничивается от предшествующей гласной по картине тех кадров, в которых она наблюдается. Следует отметить, что регистрация материала в некоторой степени зависит от уровня сигнала. Мы отмечали, например, по результатам других опытов, что чем выше уровень записи речи, тем интенсивнее пики. При этом отмечались некоторые изменения огибающей. Более того, спектр при этом превышал диапазон, отмеченный нами ранее (11500 гц). При высоком уровне записи и воспроизведения различных сочетаний звуков речи отмечались выбросы на месте согласных в полосе

частот, достигающих 16 кгц. Насколько правомерны данные аналогичных опытов, подскажут нам дальнейшие эксперименты. Особое внимание уделяли мы спектральной картине согласных звуков речи (шумных), игнорируя при этом довольно высокие пики частотных характеристик гласных, у которых в отдельных случаях отмечались на кадре срезы формантных пиков.

Теперь возьмем слово *նկիւած* (ж.). На 6, 11, 14, 16, 17, 22, 23 и 24 фильтрах отмечаем некоторые особенности фонем, составляющих данный комплекс звуков речи. Здесь можно определить характерную огибающую начального звука — сонора *ն* [n] по картине 6 фильтра (ср. аналогичную картину фильтра 5—рис. 4). На картине последующего фильтра (7) слова *նկիւած* [пехас] (рис. 5) можно наблюдать такую же форму огибающей спектральных линий. Следовательно, сонорный звук *ն* [n] имеет довольно четкую спектральную картину в полосе 142,4 ÷ 179,5 гц.

В диапазоне частот 354 ÷ 571 гц наблюдается характерная картина выбросов спектральных линий. Хотя и здесь не наблюдаются особо высокие пики амплитуды, но представлены рисунки всех звуков в последовательности, за исключением конечной аффрикаты *ծ* [с]. В начале кадра пологая линия огибающей сонорного звука *ն* [n] довольно резко возвышается на границе сочетания *նկ*- [не-], затем переходит в едва заметные линии на месте фрикативного *ի* [х]. Гласная *ւ* [а] в данной полосе частот имеет довольно плавную огибающую. Далее, на 16 фильтре (1436 ÷ 1812 гц) наблюдается резкий переход от фрикативного звука к последующей гласной *ւ* [а], причем, для последней данная полоса является формантной, так как здесь у гласной огибающая имеет рисунок узкого, острого пика.

Диапазон 22—24 фильтров — это область характерных частот конечной аффрикаты, причем в данном поясе сдвигов линий спектра в сторону более высоких частот не наблюдается. Частотная характеристика данного звука в произношении диктора-женщины аналогична картине слова *նկիւած* [пехас] в мужском произношении. Следовательно, тембральная окраска играет здесь, по-видимому, незначительную роль и не может препятствовать ходу дальнейшего, более подробного его изучения. Итак, по общей картине, после установления на вертикальной оси приближенных границ для каждого звука, уточнение которых выполняется большим количеством звуко сочетаний, получаем развернутую и подробную информацию о физической характеристике звуков речи, удовлетворяющей требованиям лингвистического анализа. Следует отметить, что испытание предложенного метода спектрального анализа согласных вовсе не ограничивается данными приведенных примеров. Авторы произвели многочисленные опыты по исследованию согласных звуков в различных позициях. Естественно, эффективность результатов по данной разработке может подтвердиться фактором неизменной точности акустической характеристики звука речи в любой позиции, с учетом специфики звукового окружения, накладывающей определенный отпечаток на физическую природу анализируемого звука. Именно этот

принцип являлся основным в наших экспериментах, в результате чего мы располагаем довольно большим иллюстративным материалом. Так, например, при исследовании аффрикаты δ [c], несомненно, недостаточны данные акустического анализа указанной фонемы в позиции абсолютно-го конца слова, как в примере, приведенном выше (см. նհիւած). Поэтому мы данную фонему рассмотрели также и в начальной, и в срединной (интервокальной) позициях, в словах $\delta ա մ$ (сам), $\delta հ \delta ա ղ$ [sicaɣ], $կ ա \delta ա ն$ [kacən], $ս ա \delta հ ի$ [sacil] и других. При этом спектральная характеристика данного звука определялась одними и теми же параметрами частот. Приведем несколько примеров. В словах $\delta ա մ$ (сам) и $\delta հ \delta ա ղ$ [sicaɣ] имеем начальный звук s (с), причем, в первом случае перед гласной $ա$ (а), а во втором—перед $հ$ [h]. Кроме того, в слове $\delta հ \delta ա ղ$ [sicaɣ] мы имеем также аффрикату s [c], но уже в позиции между двумя гласными. В описанном выше анализе, где речь идет об аффрикате s [c] для данного звука речи характерны выбросы или спектральные линии в диапазонах низких и высоких частот, при этом отмечается, что данный звук в среднем регистре характерных импульсов не имеет, что важно для анализа. Таким образом, отмечаем, что звук s (с) в зависимости от тембра речи, т. е. в зависимости от того, что данное слово произносится мужчиной или женщиной, мы имеем

$F_1 = 89,8 - 227$ для мужского голоса и соответственно

$F_1 = 142,5 - 285$ гц для женского; $F_2 = 5750 - 11500$ гц для обоих тембров, в силу чего можно, по-видимому, говорить о незначительной роли воздействия тембровой окраски голоса диктора на спектральную характеристику анализируемого звука речи в диапазоне высоких частот.

Для экспериментально-фонетических исследований звукового состава языка, для полного и всестороннего анализа, несомненно требуется максимум исследуемого материала. Любой звук речи, в качестве материала для исследовательских работ, должен рассматриваться в различных окружениях и в произношении многих дикторов для установления полной картины его физических характеристик. Такой метод количественного анализа применяется в настоящее время.

Վ. Վ. ՄԵԼԻՔԵՅԱՆ, Վ. Ն. ՀԱՅՐԱՊԵՏՅԱՆ

ՀԱՅՈՑ ԼԵԶՎԻ ԲԱՂԱԶՍՅՆՆԵՐԻ ՍՊԵԿՏՐԱԼ ՈՒՍՈՒՄՆԱՍԻՐՈՒԹՅԱՆ ՄԻՋՈՑՆԵՐԻ ՀԱՐՑԻ ՄԱՍԻՆ

Ա Մ Փ Ն Փ Ն Ն Մ

Լեզվի հնչյունների փորձառական եղանակով ուսումնասիրությունը քանակական մեթոդների կիրառման ձևերից մեկն է:

Հայերենի բաղաձայնների փորձառական ուսումնասիրությունների գործնական հետևանքը կլինի այն, որ լեզվի հնչյունների ֆիզիկական ուսումնասիրության տվյալների համաձայն հնարավոր կդառնա պատրաստել հատուկ

էլեկտրոնային սարքեր՝ լեզվի միավորների ավտոմատ ընկալման համար, իսկ հետագայում հնարավորություն կստեղծվի նաև սինթեզել բանավոր խոսքը:

Հաշվի առնելով նշված բոլոր պահանջները և ելնելով մեր հնարավորություններից, մենք լեզվի ինստիտուտի փորձառական հնչյունարանության լաբորատորիայում մշակեցինք բաղաձայնների սպեկտրալ ուսումնասիրության մի նոր մեթոդ, որը կիրառվում է գոյություն ունեցող սարքերի միջոցով, որոնց մի մասը կատարելագործել է ինժեներ Վ. Մելքոնյանը: Փորձերի ընթացքում մեր հիմնական նպատակն է եղել՝ ստանալ հնչյունների ծավալային պատկերը, որը բաղկացած է երեք գործոններից. տվյալներից ըստ տատանումների, տեղից և հզորության ամպլիտուդայի չափումներից:

Ճշգրիտ տվյալներ ապահովելու նպատակով հավաքվել է մի սարք՝ սինթրոնիզատորը, որը ապահովում է հնչյունների չափումները ըստ ժամանակի:

Ներկայացված մեթոդով մենք կատարել ենք բազմաթիվ փորձեր: Տվյալ հոդվածում զետեղված են մի շարք նյութեր փորձերի արդյունքների մասին և նկարագրված է այդ ուսումնասիրության մեթոդը: Կատարված աշխատանքը հաստատում է մշակված սպեկտրալ ուսումնասիրության միջոցի գործնական կարևորությունը: