

Санкт-Петербургский государственный университет

ЗИМИНА Светлана Владимировна

Выпускная квалификационная работа

**Акустические характеристики коммуникативной адаптации в диалоге
в сходных фонетических последовательностях**

Уровень образования: бакалавриат

Направление 45.03.02 «Лингвистика»

Основная образовательная программа СВ.5094. «Общая и прикладная
фонетика»

Профиль «Общая и прикладная фонетика»

Научный руководитель:
к.ф.н, доцент, Кафедра фонетики и
методики преподавания
иностраных языков,
Евдокимова Вера Вячеславовна

Рецензент:
к.ф.н., доцент, Кафедра
общего языкознания
имени Л.А. Вербицкой,
Риехакainen Елена Игоревна

Санкт-Петербург
2021

Содержание

Введение.....	3
Глава 1. Особенности речевой адаптации. Адаптация по акустическим характеристикам.....	5
1.1. Понятие адаптации. Определение, основные подходы.....	5
1.2. Проявление адаптации на разных уровнях языка.....	16
1.3. Исследование речевой адаптации на материале разных языков....	22
1.4. Изучение акустических спектральных характеристик речевой адаптации.....	30
1.5. Основные выводы по Главе 1.....	35
Глава 2. Исследование формантных характеристик речевой адаптации на материале диалогической речи в русском языке.....	37
2.1. Материал исследования – корпус диалогической речи SibLing...	37
2.2. Методика подбора материала из диалогов корпуса для исследования формантной адаптации в диалогах.....	43
2.3. Расчет формантных характеристик.....	47
2.4. Результаты сравнения формантных характеристик.....	50
2.5. Основные результаты и выводы по Главе 2.....	61
Заключение.....	64
Библиографический список.....	66

Введение

Актуальность исследования

В последнее время наблюдается возрастающий интерес к исследованиям в области коммуникативной адаптации собеседников. Под коммуникативной адаптацией подразумевается подстройка речевого поведения собеседников в процессе диалога.

С теоретической точки зрения существует несколько подходов к самому определению адаптации. В первую очередь важно обозначить границы самого явления, а затем искать критерии оценки. Совершенно точно можно сказать лишь о том, что это явление достаточно сложное и выходит за рамки только лингвистического подхода. Коммуникативная адаптация находится в ведении психолингвистики, так как интерес представляет не только факт ее проявления в речи людей, но и социолингвистические факторы, от которых оно может зависеть.

Определить надежные критерии для исследования наличия и степени коммуникативной подстройки людей достаточно сложно, так как тема на данный момент является новой и недостаточно хорошо изученной. Для проведения анализа выбираются разнообразные акустические характеристики. Одним из направлений является изучение варьирования формантных частот как критерия оценки наличия и степени коммуникативной адаптации.

В разных странах проводятся исследования на материале различных языков, среди которых самым популярным является английский и его региональные варианты. Также этой темой заинтересованы исследователи из Испании, Японии, Германии, Италии, Франции и других стран.

Однако, в русскоязычной научной литературе достаточно сложно найти работы, которые были бы посвящены именно анализу акустических характеристик. Данная работа является одним из экспериментальных исследований в этой сфере.

Объект исследования: особенности акустических характеристик в диалогической речи

Предмет исследования: корреляция динамики формантной структуры ударных гласных, степени и скорости коммуникативной адаптации в диалоге со степенью знакомства, социальным статусом и полом информантов

Целью данной работы стало исследование такого явления как артикуляторная адаптация собеседников в процессе диалога, основывающаяся на анализе формантных характеристик гласных. Реализации сегментов речи, обнаруженные в словах со сходными фонетическими характеристиками, сравниваются по формантным значениям

Цель работы определила следующие **задачи**:

1. Отбор материала и сегментация записей диалоговой речи на уровне слов и звуков
2. Обработка полученного материала и получение значений формант для речевых сегментов
3. Анализ полученных акустических данных и последующее установление по ним факта наличия или отсутствия коммуникативной адаптации между собеседниками

Глава 1. Особенности речевой адаптации. Адаптация по акустическим характеристикам.

1.1. Понятие адаптации. Определение, основные подходы

Основополагающей работами в изучении такого явления, как коммуникативная адаптация, можно считать 2 теории: Теорию речевой адаптации (Speech Accomodation Theory (SAT)) [57] и Теорию коммуникативной адаптации (Communication Adaptation Theory (CAT)) [34].

Теория речевой адаптации является работой, задающей основной вектор оценки этого явления. В ней появляются первые основные термины – *конвергенция* и *дивергенция*, – а также прорабатываются такие понятия, как зависимость степени проявления того или иного явления в речи собеседников от различных факторов. Более того, авторы обращают внимание на перцептивный аспект.

Суть *конвергенции* в том, что люди с большей вероятностью приближаются к особенностям речевого поведения своего собеседника, когда они хотят получить социальное одобрение с его стороны, а также тогда, когда предполагаемый результат этой подстройки требует не так много усилий. И наоборот, *дивергенция* или *сохранение своей манеры речи* проявляется тогда, когда собеседники идентифицируют себя как людей, принадлежащих к разным социальным группам и хотят сохранить дистанцию между друг другом, ярко проявляя свои особенности в межличностном общении.

Степень конвергенции в речи собеседников зависит от их речевого репертуара и факторов (внутренних и внешних), повышающих потребность в социальном одобрении. Степень дивергенции так же зависит от речевого репертуара собеседников, как и от влияния контекстуальных (ситуативных) факторов, повышающих обособленность той или иной социальной группы и желание этой группы выделиться. Также на степень дивергенции могут

влиять черты другой группы, которые бы человек из первой группы не хотел перенимать в процессе межличностного общения.

Конвергенция положительно воспринимается реципиентом в тех случаях, когда поведение человека, старающегося подстроиться под характеристики собеседника, в итоге не нарушает определенной социолингвистической дистанции между говорящими. В таком случае адресат видит в действиях своего собеседника исключительно положительные намерения. Если говорить о дивергенции, то она может рассматриваться как нечто негативное самими участниками общения. В то же время сторонними наблюдателями она будет оцениваться позитивно потому, что в ней будет наблюдаться взаимное сохранение собеседниками характеристик, присущих той или иной социальной группе. Такое поддержание социальной дистанции будет восприниматься как взаимоуважение участников общения как членов разных групп.

Существование обоих направлений изменения или сохранения характеристик говорящих также имеет функциональное обоснование. С когнитивной точки зрения, говорящий (Speaker (S)) приближается по речевым характеристикам к реципиенту (Recipient (R)) для того, чтобы облегчить взаимопонимание. Если же говорить о дивергенции, то S отдаляется по характеристикам от R для того, чтобы напомнить реципиенту о том, что они являются членами разных социальных групп, и следовательно, предотвратить присвоение несвойственных групповых черт в процессе коммуникации. Таким образом, дивергенция со стороны говорящего (S) направлена на то, чтобы побудить реципиента (R) использовать речевое поведение, которое является в данной ситуации более подходящим.

Также есть вторая сторона вопроса о функциях конвергенции/ дивергенции – это эмоциональная составляющая. С этой точки зрения конвергенция в речевых характеристиках рассматривается как то, что позволяет говорящему (S) становиться больше похожим на реципиента (R) и, следовательно, более

привлекательным. Что касается дивергенции, ее эмоциональная функция заключается в том, чтобы говорящий (S) мог подчеркнуть свои отличительные особенности, и, таким образом, обозначить свою индивидуальность (в положительном смысле).

Чуть позже теория была дополнена группой исследователей во главе с Д.Такераром [58]. Было выдвинуто предложение все же различать лингвистическую аккомодацию, подразумевая под ней само речевое поведение, и психологическую, имея в виду мотивацию и намерения говорящих сближаться или же расходиться в характеристиках.

Лингвистическая аккомодация в свою очередь подразделялась на объективную и субъективную составляющие. Имеется в виду то, что с точки зрения стороннего наблюдателя или лингвиста-исследователя может наблюдаться сдвиг характеристик говорящих в сторону, к примеру, конвергенции. В то же время субъективная оценка ситуации (ощущения самих говорящих) могут отличаться от объективных данных, и сами собеседники могут считать, что их речевое поведение отличается. Этим разделением на объективную и субъективную стороны аккомодации можно объяснить те случаи, когда наблюдается лингвистическая дивергенция в то время как собеседники намереваются сойтись и достичь объединения с психологической точки зрения.

Такое несоответствие между лингвистической и психологической адаптацией происходит во многих ролевых ситуациях, характеризующихся несоответствием статуса. К примеру, оно возможно в таких ситуациях, как взаимодействие между врачами и пациентами, профессорами и студентами или мужчинами и женщинами. В ситуациях общения с участием людей разного статуса собеседники могут вносить свой вклад в достижение общей цели с помощью различных речевых моделей. Социальные нормы в таких условиях требуют «речевой взаимозависимости» [33] вместо конвергенции. «Речевая взаимозависимость» отличается от конвергенции тем, что различия

соответствуют оптимальной социолингвистической дистанции и психологически приемлемы для обоих участников.

Таким образом, можно говорить о том, что Теория речевой адаптации (SAT) раскрыла сложность процесса коммуникации, подчеркнув важность освещения как когнитивных, так и эмоциональных процессов, лежащих в основе большого количества вербальных и невербальных моделей поведения. Эмоциональная составляющая речевой адаптации позволяет предсказывать векторы движения аккомодации говорящих (в сторону конвергенции или дивергенции) в зависимости от межличностной или межгрупповой значимости взаимодействия для них. Другими словами, она позволяет прогнозировать восприятие говорящими того, насколько на их взаимодействие влияют само- и групповая идентификация.

Имея в виду все вышесказанное об эмоциональных процессах, важно упомянуть, что они влияют не только на направление изменения речевых характеристик, но и на восприятие процесса аккомодации. В большинстве случаев конвергенция действительно вызывает у реципиента положительную реакцию, а дивергенция – отрицательную. Но также существует не менее популярное мнение – невозможно говорить об универсальности восприятия конвергенции/дивергенции абсолютно во всех ситуациях. В более позднем исследовании утверждается, что конвергенция может восприниматься положительно тогда, когда ей приписывается положительное намерение говорящего, а дивергенция оценивается отрицательно тогда, когда ей приписываются отрицательные намерения [31]. Однако это мнение не проясняет, что происходит с восприятием конвергенции в тех случаях, когда воспринимаемое намерение является отрицательным и с восприятием дивергенции в тех случаях, когда воспринимаемое намерение является положительным. К примеру, пародирование акцента (то есть конвергенция) совершенно точно будет восприниматься негативно в том случае, когда реципиент будет осознавать, что говорящий действительно насмехается над

его особенностями произношения. В то же время в случае, когда говорящий будет вынужден в виду проигрывания некоторой социальной роли подстроиться под речь реципиента, реакция последнего на такого рода подстройку будет менее негативной. Примеры подобных ситуаций подводят к выводу о том, что и конвергенция, и дивергенция могут быть расценены реципиентом как положительно, так и отрицательно.

Теория речевой адаптации делает предположение о том, что конвергенция в целом отражает желательное поведение, а дивергенция или сохранение изначальных дикторских черт – нежелательное поведение, так что появление ошибки в восприятии приводит к различным последствиям. Так, конвергенция, проявляемая членом той же группы, к которой принадлежит реципиент, воспринимается с доброжелательностью, а подстройка характеристик со стороны члена другой социальной группы приписывается к той, которая проявляется под влиянием некоторых внешних ограничивающих условий и, следовательно, рассматривается как менее желательная. Дивергенция, проявляемая членами социальной группы, отличной от группы реципиента, также ощущается как нечто враждебное, а если она исходит от членов одинаковой социальной группы – то как что-то, обусловленное внешними факторами и, следовательно, менее нежелательное.

В 1987 теория начала расширяться и переросла в Теорию коммуникативной адаптации (САТ), которая являлась расширенной и дополненной версией Теории речевой адаптации (SAT). Она уже предполагала, что при оценке направления движения изменения речевых характеристик, что должна быть опора на историю межгрупповых отношений, статус групп, внутригрупповые связи, стабильность и закономерность внутригрупповых отношений, социальные нормы для межгрупповых контактов, а также культурные ценности. Такое расширение сделало теорию слишком сложной. Уже в 1998 С.Галуа и Г. Джайлз писали о том, что вся теория не может быть проверена на практике одновременно, поэтому исследователям, использующим ее

положения, стоит развивать мини-теории, которые бы соответствовали определенному материалу, но с опорой на большую Теорию коммуникативной адаптации [30].

Более того, Стрит и Джайлз [57] обозначили то, что конвергенция может быть как полной, так и частичной. Так, человек, говорящий со скоростью 100 слов в минуту, может постараться подстроиться под своего собеседника, произносящего 200 слов в минуту. В случае, когда скорость речи диктора, который изначально говорил медленнее, повысится до 200 слов в минуту, можно будет говорить о том, что произошла полная конвергенция. Если же он станет произносить, к примеру, 150 слов в минуту, то конвергенция будет считаться частичной.

Еще один подход к речевой адаптации заключается в том, что она бывает унимодальной и мультимодальной [27]. Суть унимодальной заключается в том, что люди в процессе коммуникации меняют только один аспект общения – к примеру, акцент. Мультимодальная адаптация отличается тем, что смещению подвергается сразу несколько аспектов – акцент, жест, поза. Проиллюстрировать оба вида может исследование, которое провели Р.Бурхис и Г.Джайлз [35]. Когда носители уэльского акцента сосредотачивались на своем валлийском акценте, это можно считать проявлением унимодальной адаптации. В том случае, когда обращается внимание уже не только на акцент, но и на фразы, новую лексику, можно говорить о мультимодальной адаптации. Также собеседники могут проявлять как конвергенцию, так и конвергенцию в рамках одного и того же диалога. К примеру, они могут сходиться в одних темах и расходиться в других. В работе Ф. Билоуса и Р. Краусса [23] как раз рассматривается случай такого проявления сложного вида конвергенции: женщины в диалогах с мужчинами сходятся по таким характеристикам, как паузация, длина фраз и перебивки, но в то же время могут расходиться с ними в смехе.

Еще один из подходов касается направления конвергенции или дивергенции. Г. Джайлз и П. Поусленд в своем труде [37] определили, что вариация в речевом поведении обусловлена социальными причинами. «Конвергенцией вверх» считается приближение к стилю речи, который является более престижным. Наоборот, «конвергенция вниз» - смещение речевых характеристик в сторону менее авторитетного стиля.

Также конвергенция может проявляться как в долгосрочной, так и в краткосрочной перспективе. Эксперимент, проведенный Д.Пардо [49], действительно доказывает существование долгосрочной конвергенции. Исследовательница отмечает, что взаимная подстройка соседей по комнате только возрастает с течением времени (учебного года).

Важно отметить, что есть и негативные варианты развития коммуникативной подстройки. Существует такое понятие, как чрезмерная адаптация. Одним из ее проявлений можно считать то, что представители молодого поколения в разговоре со старшими, в попытках подстроиться под их речь, делают слишком много повторов, преувеличенно имитируют интонацию, становятся излишне вежливыми и говорят громче [56]. Также возможна и недостаточная адаптация, которая проявляется в том, что собеседники не достигают достаточного уровня сближения с точки зрения речевого поведения [36]. Одним из примеров можно считать неспособность выбрать тему для того, чтобы ответить собеседнику [38]. Для иллюстрации можно также рассмотреть отношения молодого и старшего поколений. Диалог молодых людей с пожилыми зачастую не складывается потому, что последние предпочитают говорить о своих недугах или же выбирают темы, которые недостаточно привлекательны для молодых людей [39]. Более того, недостаточная адаптация может доходить даже до намеренной дезадаптации. Д. Гасиорек и Г.Джайлс определяют ее как отсутствие должной коммуникативной подстройки со стороны хотя бы одного говорящего [32]. Отказ от адаптации может быть как преднамеренным, так и происходить без

какого либо влияния решения говорящего. Намеренное расхождение в речевом поведении может зависеть от различий в культурных ценностях [26], а также может являться результатом игнорирования потребностей или общих знаний участника(ов) разговора [32]. В такой ситуации различные компоненты (выбор темы, лексики, грамматические правила, жестикация), возможно, преднамеренно употребляются неправильно, что приводит к тому, что собеседники не понимают намерений друг друга. Также Д. Хьюитт, Б. Уотсон и С. Галлуа делают предположение о том, что непонимание и неудовлетворения – это то, к чему приводит отсутствия между собеседниками [40].

Работая в направлении изучения подстройки характеристик собеседников в процессе коммуникации, Левитан и Хиршберг подробнее описали степени согласованности [45]. Так, авторы выделяют 3 градации: сближение (proximity), конвергенцию (convergence) и синхронизацию (synchrony) (рис.1.1).

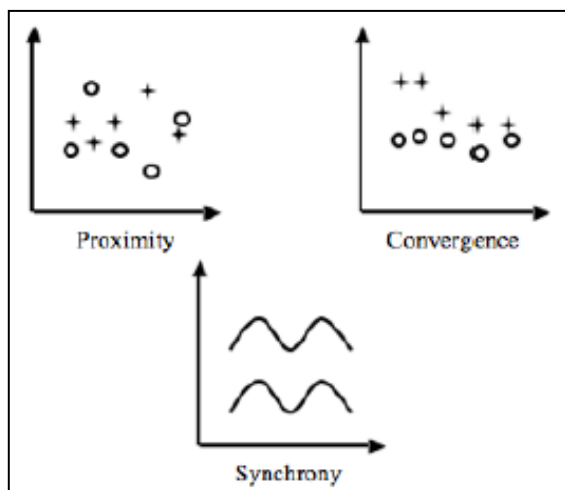


Рис.1.1. Различные виды речевой адаптации: на оси x отмечено время, на оси y – величины определенной характеристики. В виде плюсов и окружностей изображены значения для 2 разных собеседников [45].

Каждый из типов приближения речевых характеристик рассчитывался особым способом. Сближение проверялось следующим образом: речь

отдельного диктора сравнивалась с речью партнера (по одному из выбранных измерений) и со всеми другими, с кем этот диктор не разговаривал в ходе эксперимента. Также одним из способов расчета сближения является сравнение записей одного и того же информанта с самим собой, но в разных диалогах..

Расчет конвергенции проводился на материале 2 отрезков речи информантов: в начале разговора и в конце. В ходе исследования было обнаружено, что расхождения по всем параметрам между собеседниками значительно меньше во второй части диалога нежели в первой.

Что касается синхронизации, о ней можно судить в том случае, когда характеристики двух дикторов согласуются друг с другом, но не совпадают. Таким образом, синхронизация возможна без сближения.

Для проведения анализ учеными были выбраны следующие параметры: интенсивность – максимальная в каждой реплике (отрезке, в котором не было пауз) и средняя; просодические характеристики (максимальный и средний уровни частоты основного тона); качество голоса: дрожание амплитуды периодов основной частоты (*shimmer*), дрожание длины периода основной частоты (*jitter*) и соотношение шума и голоса (*HNR*); скорость речи, вычисляемая по словарю автоматически, единица измерения – слоги в секунду.

Более того, были учтены гендерные различия дикторов. Предварительно была проведена нормализация для всех характеристик, а именно рассчитана средняя величина, соответствующая гендерной группе, а также ее стандартное отклонение.

В ходе исследования была рассчитана степень речевой адаптации с учетом приведенных выше 3 измерений. По результатам анализа было подтверждено сближение собеседников, так как различий в паре было всегда меньше, чем между каждым отдельно взятым диктором и другими, с которыми он не

вступал в диалог. В большей степени адаптация затронула среднюю и максимальную интенсивность, максимальные значения частоты основного тона. Меньше дикторы адаптировались по такой характеристике, как дрожание амплитуды ОТ и по скорости речи.

Что касается схожести диктора с самим собой в другом диалоге нежели со своим партнером, ее можно было проследить, опираясь на следующие параметры: скорость речи, дрожание амплитуды периодов основной частоты (shimmer), дрожание длины периода основной частоты (jitter), соотношение шума и голоса (HNR), средняя частота основного тона. Однако, проанализировав максимальную и среднюю интенсивность, можно говорить о том, что по этим характеристикам собеседники ближе друг к другу, нежели к самим себе в других диалогах. Это может свидетельствовать о том, что люди для подстройки под речь друг друга меняют свою нормальную громкость речи.

Для того, чтобы достигнуть согласованности с партнером, не меняя индивидуальных характеристик, участники разговора подстраиваются по таким параметрам, как мерцание и скорость речи. Это единственные характеристики, по которым в итоге дикторы были больше похожи на самих себя, нежели на своих собеседников, и больше схожи с собеседниками, чем на других информантов.

Также расчет параметров проводился на границах реплик. В этих местах сближение оказалось еще более ярко выраженным, чем в рамках всего диалога. Более того, оно проявлялось и в тех параметрах, которые не были сильно близки в процессе всего разговора. Следовательно, можно сделать вывод о том, что взаимная подстройка дикторов к речи друг друга может затрагивать разные параметры в разное время. То же самое можно сказать и о увеличении сближения во второй части беседы по сравнению с первой. К примеру, такой параметр, как качество голоса (HNR), схож у дикторов только

во второй части диалога. Это можно объяснить тем, что этому параметру, в отличие от той же интенсивности, нужно больше времени для изменения.

1.2. Проявление адаптации на разных уровнях языка

Рассуждая о том, как речевая адаптация может затрагивать разные уровни языковой системы, стоит упомянуть о работе М. Пикеринга и С. Гэррода [52]. Психологи предлагают свой вариант модели подстройки в диалоге, которая бы учитывала взаимодействие собеседников, затрагивающее все уровни языка (рис.1.2).

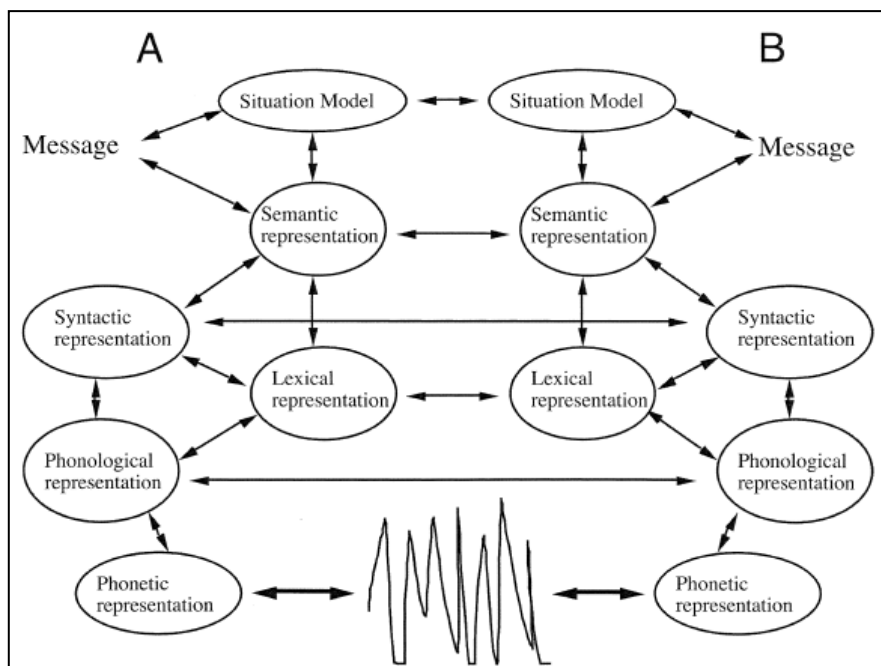


Рис.1.2. А и В представляют собой двух собеседников в этой схеме этапов восприятия и производства речи в соответствии с моделью подстройки в диалоге [52]. Модель отражает основные уровни системы и связи между ними, а также связи, формирующиеся в процессе коммуникации А и В.

Данная модель предполагает, что успешный диалог подразумевает развитие согласующихся образов в сознании собеседников. Это происходит благодаря подготовительным процессам на каждом уровне языкового представления путем фильтрации информации. Таким образом выравнивание на одном уровне усиливает выравнивание на других уровнях.

На рис.1.2. можно видеть процесс подстройки на разных уровнях, изображенный в достаточно абстрактном, общем виде. Модель предполагает,

что в процессе разговора в сознании собеседников происходит соотнесение отдельных компонентов на каждом уровне языка. Важно отметить, что выстраиваются не только связи между языковыми уровнями каждого участника разговора, но и общие для обоих собеседников.

Ученые обращают внимание на то, что традиционно все изменения в речи рассматривались с точки зрения монологической речи. М. Пикеринг и С. Гэррод говорят о том, что существовавшая ранее модель «автономной передачи информации» является ошибочной, так как она не предполагает никакой связи между собеседниками. Во-вторых, авторы работы рассматривают эти связи как те, которые соответствуют каналам, на которых происходит подготовка информации для следующих уровней. Более того, делается акцент на двунаправленный характер связей: это означает то, что должно существовать равновесие между процессами производства речи и ее восприятием.

Недостаток автономной модели передачи информации в том, что между говорящим и слушающим речепроизводство и понимание речи происходят изолированно друг от друга. Тот, кто говорит, формулирует высказывание на основе своего представления ситуации. Строго говоря, идея, сформировавшаяся в сознании, нелингвистическая, преобразуется в ряд лингвистических явлений. Самыми первыми появляются синтаксические единицы, а фонологические – в последнюю очередь. Конечное лингвистическое представление преобразуется в артикуляторную программу, итогом которой является звук или жестикуляция. Существуют, соответственно, промежуточные этапы, которые являются значимыми только для говорящего, то есть остающиеся внутренними. Следовательно, нет никаких оснований, по которым бы слушателю должны были быть важны процессы, происходящие в промежуточные стадии.

Слушающий же в свою очередь декодирует сигнал, преобразуя его в своем сознании в лингвистические репрезентации на разных уровнях. Процесс

происходит до тех пор, пока исходное сообщение не будет до конца восстановлено. После этого слушающий делает определенные выводы, то есть судит о сказанном, исходя из своего видения ситуации. Таким образом, с точки зрения обработки сигнала, говорящий и слушающий действуют изолированно. Единственное, что остается общего между собеседниками – сообщения, передаваемые ими друг другу.

Каждый акт передачи информации рассматривается как дискретная стадия, на которой происходит кодирование говорящим своего сообщения в звуковые единицы, а затем – последующее декодирование сигнала слушающим. Более того, нет даже оснований для того, чтобы предполагать ни то, что уровни собеседников будут совпадать, ни то, что уровни восприятия и производства речи будут одинаковыми.

Таким образом, модель автономной передачи информации является неподходящей для диалога, так как в диалоге наблюдается тесная связь между речепроизводством и восприятием. При формулировании своих высказываний говорящий опирается на то, что только что было ему сказано. Также и слушающий оказывается ограничен тем, что перед этим произнес говорящий. Диалог является результатом совместной деятельности собеседников, в процессе которой производство речи и ее восприятие тесно переплетаются, в свою очередь согласующиеся на нескольких уровнях. Так, каждый уровень представления информации оказывается тем временем вовлеченным в процесс коммуникации. В результате эти промежуточные стадии сохраняются; их существование хотя бы подразумевается в усовершенствованной модели. Именно потому, что подстройка на одном уровне приводит к подстройке и на других тоже, собеседники приходят к согласованию своих взглядов на ситуацию и, следовательно, понимают друг друга. Такой итог беседы возможен только при реализации интерактивной модели подстройки, но не при реализации модели автономной передачи информации.

Интерактивная модель подстройки предполагает, что процессы, используемые для производства речи, должны опираться на те, что происходят при ее восприятии. Это равновесие означает то, что репрезентация, выстроенная в целях понимания, может впоследствии использоваться для речепроизводства и наоборот. Так, эта модель объясняет, почему люди, могут, к примеру, безошибочно дополнять высказывания друг друга на уровне синтаксиса, семантики и фонологии.

К примеру, ответ на вопрос «Is Jack in town?», вне зависимости от того, к какому слову поставлен вопрос, потребует синтаксического параллелизма. Это значит, что при ответе нельзя будет использовать не «he», а «him». Предложение для уточнения, построенное по типу «Jack?», уже будет задействовать параллелизм на фонологическом или, возможно, на фонетическом уровне.

Следуя за исследованием М. Пикеринга и С. Гэррода [52], группа ученых, А. Ненкова, А. Гравано и Д. Хиршберг, предлагают изучение подстройки дикторов с опорой на сверхчастотные слова [48]. Исследование проводилось на материале телефонных разговоров The Switchboard Corpus [39]. Дикторы – носители различных вариантов Американского английского. Аннотация корпуса интересна тем, что в ней была отмечена степень естественности диалогов по шкале Ликерта: от 1 (звучит очень естественно) до 5 (звучит совсем неестественно). Высокочастотные слова из всего корпуса использовались в качестве средства, на которое можно было бы опираться при ранжировании диалогов. Гипотеза исследователей заключается в том, что такая оценка является более надежной и имеет более общий характер, чем попытки изучить адаптацию по часто встречающимся словам в конкретном диалоге.

Была разработана метрика для подсчета адаптации. Она рассчитывалась как разница для 2 дикторов между отношениями количества реализаций какого-то конкретного слова ко всем словам, произнесенным диктором. Для

проведения анализа выбрано 25 самых частотных слов. Согласно результатам исследования, такую метрику можно считать надежным показателем естественности диалога (с точки зрения перцепции).

Эта метрика позволила не только вычислить степень адаптации, но и согласованность диалога: она шла в негативной корреляции с перебивками собеседников, что является очень важным результатом для разработки диалоговых систем (SDS). Также от этого зависит успешность выполнения задания и правильная смена реплик в процессе разговора.

Также о том, что адаптация на семантическом и лексическом уровне выгодна для построения диалога как говорящему, так и слушающему, говорит и группа ученых во главе с Х. Брэниган [24]. Слушатель извлекает выгоду из подстройки потому, что она повышает вероятность правильного понимания смысла услышанного. Говорящему же она приносит пользу в том плане, что она снижает нагрузку на его мозг: уже обработав информацию определенным образом, он определяет для себя семантическое поле, которое будет использоваться для речепроизводства. Тому, кто говорит, не нужно заново выстраивать модели для того, чтобы что-то сказать, достаточно опоры на то, что уже было сказано собеседником. Эти преимущества (с функциональной точки зрения) ведут к тому, что высказывания собеседников будут семантически и лексически соответствовать предыдущим высказываниям.

Эксперимент был построен на карточной игре, которая предполагает описание карточек различными способами. Исследователи отслеживали динамику употребления 2 типов конструкций, используемых дикторами при объяснениях. Было замечено, что говорящие чаще реализуют ту же грамматическую форму, которую только что услышали от собеседника. Исследователи исключали возможность влияния дискурса, употребления одних и тех же глаголов и слов, так как это никак бы не объясняло, все же, использование одинаковых синтаксических конструкций. Обращается внимание на то, что координация происходит именно по форме, а не по

значению. Более того, результаты выдающимися еще и потому, что использование одной и той же формы совсем не являлось обязательным для достижения успеха в процессе коммуникации.

Также эти данные подтверждают то, что в процессе речевого производства собеседники восприимчивы к общей коммуникативной ситуации, в частности – к языковому поведению других участников разговора. Авторы акцентируют внимание на том, что они не говорят того, что собеседники осознанно подстраивали речевое поведение друг под друга, чтобы соответствовать друг другу. То, что в рамках данной работы называется координацией языкового поведения, может возникать без какого бы то ни было намерения со стороны участников диалога.

1.3. Исследование речевой адаптации на материале разных языков

В зарубежной литературе можно найти достаточно много работ, посвященных теме вычисления степени коммуникативной адаптации. Самыми популярными среди них можно считать исследования, которые проводились на материале английского языка, в том числе его диалектов.

Одна из самых часто встречающихся методик обнаружения факта наличия или отсутствия речевой адаптации – перцептивный эксперимент, а именно АХВ сравнение. Такой метод используется Д. Пардо в работе, посвященной анализу фонетической конвергенции в диалоге [51]. Сначала проводились записи до выполнения заданий, затем, через 1-2 недели дикторы разговаривали в тех же парах, выполняя задания с маршрутами, а после задания отдельно записывались образцы речи. Эксперимент строился таким образом, что слушателю предоставлялись 3 образца фразы, включающей в себя название одной из локаций (образцы соответствовали этапам записи, описанным ранее). В ходе анализа данных было обнаружено, что сближение в фонетических характеристиках дикторов, причем достаточно сильное. Несмотря на то, что участники эксперимента не были знакомы до взаимодействия, в ходе разговора произошла коммуникативная адаптация, которая усиливалась в ходе беседы, а также сохранялась и после завершения выполнения общего задания (игры). Что касается факторов, влияющих на конвергенцию: автором исследования в первую очередь отмечается зависимость от пола диктора. В целом было выявлено, что адаптация между мужчинами сильнее, чем между женщинами. Также степень адаптации зависела и от роли говорящего: объясняющие адаптировались к тем, кому объясняли. В женских парах те дикторы, которым объясняли, не демонстрировали адаптации к объясняющим, в отличие от мужских пар, в которых ситуация противоположная. Таким образом, это исследование устанавливает факт наличия относительно быстрого процесса фонетической

конвергенции между собеседниками, зависящего от роли и пола говорящего и сохраняющегося за пределами разговора.

Д.Пардо отмечает, что результаты исследования идут в разрез с предположениями, выдвигаемыми в более ранних работах другими исследователями. В данном контексте упоминается работа Л. Нэми, в которой делается акцент на роли фактора пола в степени коммуникативной адаптации [47]. На материале заданий, построенных на повторе ранее сказанных слов (shadowing task), было обнаружено, что женщины лучше адаптируются, нежели мужчины, и больше к мужчинам, чем к женщинам. Более того, слушателями-женщинами более активно отмечался факт адаптации, чем мужчинами.

Также в англоязычной среде довольно популярны исследования, подходящие к коммуникативной адаптации со стороны диалектного варьирования. Интерес представляет работа, посвященная американскому и индийскому вариантам английского [44]. Гипотеза авторов состояла в том, что различия этих двух региолектов заключаются в длине ударных слогов в стопах. Американский вариант предполагает удлинение ударного слога в односложной стопе, контрастирующее с его отсутствием в многосложных. Что же касается индийского, в нем такой разницы не наблюдается. Длительность стопы в индийском варианте зависит от количества слогов в ней, в то время как в американском варианте такого не наблюдается. Таким образом, дивергенцию или конвергенцию можно проследить, опираясь на данные прочтения текста: в первый раз они читают текст в одиночку, второй – вместе с диктором, говорящим на другом диалекте.

В ходе эксперимента было установлено, что при синхронном прочтении текста слоги, находящиеся под ударением, действительно длиннее в односложных стопах; как в индийском, так и в американском вариантах. Также авторами было замечено то, что темп речи стал более замедленным. В речи одного из дикторов-индейцев во время синхронного чтения текста

совместно с диктором-американцем наблюдалась редукция по длительности ударного слога, которая стала сильнее относительно одиночного чтения. Конвергенция в случае этого отдельно взятого информанта проявляется в том, что речь по ритму приближается к тактосчитающему типу, которому свойственно, что ударные слоги произносятся примерно через одинаковые временные периоды [25].

Что касается американского варианта, для него было свойственно удлинение стопы, коррелирующее с увеличением количества слогов, входящих в ее состав. Более того, стопа стала длиннее относительно варианта при одиночном прочтении, когда дикторы, говорящие на двух разных диалектах, читали текст вместе. Следовательно, можно сделать вывод о том, что во время взаимодействия носителя американского диалекта с носителем индийского в речи первого начали проявляться черты, свойственные для слогосчитающего языка. Это такой тип, для которого характерно то, что слоги реализуются в равных временных интервалах, и, следовательно, их длительность примерно одинаковая [25]. Таким образом, можно судить о проявившейся конвергенции.

Относительно темпа речи авторами также отмечалась конвергенция между дикторами: носители американского варианта по темпу становились ближе к носителям индийского варианта английского. Однако, стоит заметить, что диктор, говорящий на индийском варианте, долгое время проживал в Америке, что в теории могло повлиять на результаты, так как за это время он мог перенять ритмическую структуру американского диалекта.

Исследователями из Германии была предпринята попытка проследить корреляцию между приятностью, привлекательностью голоса и его акустическими характеристикам [55]. Эксперимент проводился с использованием немецкого корпуса разговорной речи, все диалоги были женскими. Выбор темы был абсолютно свободным. Записи проводились в двух разных условиях: когда дикторы были изолированы друг от друга

звуконепроницаемой стеной и не видели друг друга и когда собеседники могли видеть друг друга через прозрачную перегородку

После каждого диалога информанты заполняли анкеты, в которых делились своими впечатлениями о собеседнике и о диалоге в целом. Каждый из параметров социальной привлекательности, то есть то, насколько человек приятный, дружелюбный, грамотный/ компетентный и уверенный в себе, было предложено оценить от 1 до 5 по шкале Ликерта в диапазоне от «да» до «нет». Для проведения эксперимента были выбраны долгие гласные, длина которых составляла не менее 0,1 секунды в целях получения более надежных результатов. В качестве фонетических параметров были выбраны медиана частоты основного тона, ее амплитуда падений и подъемов; длительность слогов, качество голоса (мерцание, дрожания, соотношение шума и гармоник (HNR), интенсивность, а также первая и вторая форманты.

В ходе эксперимента было обнаружено, что дикторы, которые были оценены как уверенные в себе, в большей степени сходятся по параметру амплитуды подъема частоты основного тона и по длительности слогов, а также с точки зрения расстояния между векторами слов и их частичного совпадения. Более того, дикторы, получившие такую оценку, сходились еще и в частотных значениях первой форманты. Также в целом дикторы оценивались как более уверенные в себе тогда, когда собеседник их видел через прозрачную перегородку.

Что касается грамотности или общей компетентности, стоит отметить, что этот фактор оказался не столь значимым для конвергенции, так как был обнаружен только один фактор, коррелирующий с этой оценкой. Однако, было замечено, что есть и другие факторы, связанные с оценкой компетентности диктора и не связанные с конвергенцией: дикторы, использующие в речи больше слов с позитивным значением, но меньше смеющиеся, оценивались как более компетентные в тех диалогах, когда дикторы не видели друг друга. Тем не менее, ситуация с влиянием этих

факторов обратная в том случае, когда собеседники находились друг перед другом. Более того, дикторы, в голосе которых было зарегистрировано меньше дрожания амплитуды периода ОТ, оценивались как более компетентные в общении.

Если говорить о дружелюбии, то с этой оценкой коррелирует только один статистически значимый фактор. Дикторы, которые обладают наименьшим диапазоном скачков интенсивности, оцениваются как более дружелюбные.

Также было обнаружено, что за восприятие человека как привлекательного, отвечают 2 фактора, которые в свою очередь связаны с конвергенцией. Дикторы, имеющие меньшее относительное расстояние (больше сходятся в слове, которое имеет наибольшую значимость), воспринимаются как более привлекательные. Интересно, что сходимость с точки зрения максимального значения частоты основного тона, также актуально для конвергенции, хоть только в случае, когда дикторы могли видеть друг друга. Однако, стоит отметить, что сложно проследить, как оценка привлекательности зависит от направления адаптации частотного пика: наблюдалось, что дикторы, которые меньше подстраивают под речь собеседника свое максимальное значение частоты основного тона, вызывают большую симпатию.

Темой коммуникативной адаптации заинтересовались и ученые из Франции, которые подошли к этой теме с точки зрения фонетики и фонологии [22]. Эксперимент Ж.Байи и А. Лелонг построен на основе игры. Собеседникам была предложена так называемая игра в домино: участникам нужно было выбрать из списка слов такое, которое бы начиналось с того же слога, который был ранее произнесен собеседником. Таким образом, можно было получить большое количество образцов гласных. Впоследствии сравнивались результаты незнакомцев, говоривших впервые и друзей, знавших друг друга и работавших друг с другом на протяжении нескольких лет. Важно отметить, что проводилось 3 серии экспериментов: когда незнакомцы и друзья

находились в разных комнатах и общались друг с другом через аппаратуру, а в третьем случае – друзья сидели лицом друг к другу.

Для исследования использовалась полуавтоматизированная технология отслеживания формантных частот. Исследования показали, что в случае с незнакомцами обнаруживалась лишь небольшая конвергенция, в то время как друзья сильнее адаптировались к речи друг друга. Более того, учеными отмечается, что конвергенция также зависит от типа гласного: огубленные гласные в большей степени подвергаются этому явлению. Дикторы в разной степени смещают гласные, особенно это касается гласных среднего ряда.

Что касается смещений, наблюдаемых в просодике, учеными отмечается лишь незначительное влияние коммуникации на этот фактор. Однако, было замечено, что конвергенция достаточно сильно проявляется в длительности гласных начальных слогов.

В университете Бари (Италия) тоже занимаются вопросами коммуникативной адаптации, подходя к вопросу с точки зрения изменений в просодике и их корреляции с отношением к говорящему [54]. Каждый из диалогов, в ходе которых дикторы играли в игру, был разделен на 2 части для того, чтобы отследить динамику параметров в первой и второй части записи. В качестве характеристик для наблюдения за просодической конвергенцией, были выбраны частотный диапазон основного тона (от минимального до максимального), средний уровень частоты основного тона, громкость и скорость артикуляции.

В ходе эксперимента действительно были обнаружены данные, указывающие на конвергенцию или дивергенцию в четырех из пяти диалогов (в рамках этого исследования авторы считают дивергенцию особым ходом развития адаптации). Так, в одном из диалогов дикторы сблизилась по громкости голоса во второй части диалога. В другом диалоге собеседники адаптировались друг к другу, но путем расхождения в скорости артикуляции

во второй части записи. В третьем и четвертом диалогах замечены оба типа проявления согласованности в диалоге. Участники третьей записи сошлись в скорости артикуляции и отдалились по признаку громкости. Дикторы в четвертой записи приблизились по диапазонам частоты основного тона и разошлись в средней частоте основного тона.

Также исследователями отмечалось, что симпатия (или один из типов дружелюбности) находится в некоторой зависимости от просодических характеристик. Измерялась степень асимметрии в степени сходимости/расхождения по параметрам для каждого диктора в паре в обеих частях диалога. В результате те дикторы, абсолютные значения для которых были больше (значит, они «больше сходились») или меньше (значит, «меньше расходились») по отношению к своим собеседникам, получали больше баллов по шкале симпатии в паре.

На материале некоторых других языков (испанский [53], японский [43] и др.) также проводились исследования в области речевой адаптации. Однако, до недавнего времени в русском языке таких работ не было. Те исследования, которые были проведены, больше похожи на обзорные в данной тематике [20], либо же углублялись в смежные области – к примеру, в фоноскопию [11]. В последней работе больше делается упор на психологическую сторону вопроса адаптации, нежели на отдельно взятые акустические характеристики.

С появлением нового корпуса SibLing, разработанного специально в целях изучения коммуникативной адаптации, увеличилось и количество работ, связанных с этой темой [41]. Стоит отметить работу, посвященную акустическим характеристикам адаптации в диалоге, проведенную на материале русского языка [46]. Для анализа были выбраны такие характеристики, как частота основного тона, энергия, соотношение шума и голоса (HNR), падение спектра от низких к высоким частотам, дрожание амплитуды периода OT и дрожание длины периода OT. Степень адаптации

оценивалась в трех измерениях: сближение, конвергенция и совпадение/синхрония [45].

Одинаковые значения для сближения в характеристиках наблюдаются только в 44 случаях из 70: 33 незначительных (пустые ячейки), 10 отрицательных и 1 случай полного сближения. Все же остальные являются примерами, в которых в течение диалога происходило варьирование по параметрам. В большем количестве случаев сближение по параметрам отрицательное и проявляется во всех акустических характеристиках. Чаще всего сближение происходит по частоте основного тона.

Что касается конвергенции, стоит отметить, что она проявлялась только в максимальном значении падения спектра от низких к высоким частотам. В остальном же преобладает отрицательная конвергенция, хотя замечены и случаи положительной. Отрицательная чаще всего проявляется в энергии сигнала, следовательно, в плане громкости дикторы чаще всего предпочитают проявлять дивергенцию. Однако, конвергенция по максимальной частоте основного тона, рассчитанная в пределах целой реплики, является положительной. Возможно, расчет по межпаузальным отрезкам речи не подходит конкретно для данной характеристики, так как они являются слишком короткими для того, чтобы на них проявилась динамика.

Если говорить о синхронии, то исследователями отмечается большое количество положительных случаев. Чаще всего она находит отражение в таких параметрах, как энергия сигнала и характеристиках, связанных с качеством голоса.

1.4. Изучение акустических спектральных характеристик

Разработанная Г.Фантом теория речеобразования и выделение в речевом тракте двух основных частей – голосового источника и артикуляторного аппарата – позволила в дальнейшем проводить расчеты, касающиеся формантных характеристик звуков[8].

По мере развития вычислительной техники появлялись и новые возможности для акустического анализа речевого сигнала. Наряду с идентификацией источника звука, изучением супрасегментных характеристик сигнала и машинным синтезом речи стало возможным проводить анализ вариаций частот формант. Задача определения траекторий движения формант является той, без которой сложно представить обработку речевого сигнала. Ее выполнение необходимо для идентификации звуков, изучения супрасегментных единиц, настройки алгоритмов синтеза речи. Именно поэтому необходимости вычисления динамики частот формант в настоящее время уделяется большое внимание.

В целом, методы исследования можно разделить на два типа: автоматизированные и неавтоматизированные. К последним относится автоматическая обработка окном, которая проводится на сегментах. Такой метод является эффективным, поскольку он позволяет получить достаточно много данных о структуре сигнала. Однако для высокой степени точности данных и минимизации ошибок выделение формант происходит вручную, по спектрограммам. Логично предположить, что при обработке окном достаточно большого объема данных увеличивается и время, затраченное на тщательный анализ. Поэтому становится важным перейти к автоматизации процесса.

Если же говорить о методах, которые относятся ко второму типу, а именно – об автоматизированных, то следует учесть, что они также имеют свои

минусы, как и методы первого типа. Автоматизированные методы обработки речи в свою очередь также подразделяются на несколько групп:

1. Первая группа методов не использует фонетического подхода и не учитывает различные физиологические особенности речевого тракта. Однако их применение позволяет строить передаточные функции и осуществлять переход от временного сигнала к его частотным характеристикам. К тому же, становится возможным отслеживание динамики формант во времени.

Средством является «вейвлет-преобразование» [8], которое позволяет совершить переход из временного представления сигнала в частотно-временное. Использование именно этого преобразования предоставляет информацию о появлении отдельных частот в сигнале, а также о вариации формант во времени. При корректировке результатов может быть задействована заранее подготовленная база данных, содержащих информацию о формантах гласных [6].

2. Построение вектора состояния априорно заданной динамической системы или же речевого тракта

Средство – фильтр Калмана. Это фильтр, который позволяет сглаживать временные скачки формантных частот. Он сглаживает влияние шума, используя текущие замеры и уже считанную ранее информацию [6].

3. Метод, использующий заданную структуру передаточной функции речевого тракта. Также есть информация, что сигнал делится на источник и фильтр, и параметры последнего задаются заранее.

Средство – LPC-преобразование (линейное предсказательное кодирование). LPC-преобразование используется в программном обеспечении PRAAT.

Некоторые исследователи добавляют сглаживание скачков формант, применяя алгоритм быстрой сходимости Ньютона-Рафсона с итерацией с оптимальным шагом [6].

4. Использование различных преобразований для определения формант при заранее заданной структуре речевого тракта
Средства: LPC-преобразование, SWLP-преобразование, MDVR (минимизация дисперсии искажения реакции) [6].
5. Выбор для каждого сегмента самой ровной формантной траектории. Вычисления проводятся на коротких отрезках речи. Одно из главных достоинств метода – полная автоматизация процесса
Средства: алгоритмы программы PRAAT [6].
6. Отслеживание формант, адаптированное под вокальную речь.
Средства: форманты оцениваются методами гибридного линейного прогнозирования (HLP), синхронизированного с частотой основного тона [6].
7. Автоматизированные алгоритмы выделения формант и поиска выровненных фрагментов (таких, на которых частоты F1 и F2 совпадают). Такие алгоритмы помогают делать выводы о том, совпадают дикторы или нет. Так же, как и метод предыдущего пункта, можно считать методом с узкой специализацией (как пример – криминалистика, фоноскопическая экспертиза).
Средства: программа, написанная с использованием пакета Matlab 2018a, в дальнейшем – нейронные сети [3].
8. Математический аппарат частотных методов исследования динамических систем как метод построения формантных картин гласных. Данным методом можно получить любую степень детализации формантных частот или амплитудных частотных характеристик (в зависимости от размера окна обработки), а также доступна полная автоматизация обработки речевого сигнала.
Средства: модель Лильенкрантца-Фанта (учитывается 2 звена речевого тракта и описывается динамика каждого из отделов) [6].
9. Метод нулей сигнала. Основан на анализе распределения длительностей интервалов между нулями сигнала. Является более помехоустойчивым,

нежели те, в которых используется LPC-преобразование и спектральный анализ. Однако он предполагает, что в данной частотной полосе присутствуют колебания только одной форманты, а потому выбор полос частот для анализа является существенным [18].

10. Кепстральный анализ. Кепстр – математическое преобразование, заключающееся в том, что спектральному преобразованию подвергается спектр функции. Далее в полученном кепстре выявляются гармоники. Средство: получение кепстра (обратное преобразование Фурье логарифма спектра мощности сигнала) [12].
11. Анализ 2 видов сигнала: прямо над голосовыми связками и того, который прошел через резонаторные полости. Анализ показал, что частоты, похожие на форманты, могут быть зафиксированы и прямо над связками. Это может быть вызвано тем, что акустическая энергия речевого тракта все же вернулась в надсвязочную область. Средство: спектральный анализ 2 сигналов [29].

Стоит отметить, что существует огромное множество методов проведения формантного анализа, начиная теми, что проводятся вручную, и заканчивая теми, которые автоматизированы и задействуют нейронные сети.

Метод, используемый в рамках данной работы – построение формантных картин. Его можно назвать частично автоматизированным, так как сначала используется программа, рассчитывающая формантные значения в нескольких точках, а затем по полученным усредненным данным выстраивается график. Более подробное поэтапное описание данного метода можно найти в Главе 2.

Стоит отметить удобство использования этого метода именно для анализа речевой адаптации. Формантная картина гласного может измениться от малейшего смещения органов артикуляции. Соответственно, гласные являются очень изменчивыми как акустически, так и перцептивно. На этом

основании можно считать, что гласные являются хорошим материалом для изучения конвергенции [50].

Как и в статье Д. Пардо, зачастую результаты построения формантных картин уточняют, рассчитывая дополнительно Евклидово расстояние.

Так, в одном из исследований на тему адаптации М.Бабель делает выводы с опорой именно на эти два средства [21]. В ходе этого эксперимента исследовательница выясняла, как сдвигаются форманты в процессе выполнения заданий. В данном случае использовалась методика теневого повтора *shadowing* (задача диктора: повторить за человеком, чей голос он слышит в наушниках). В одном случае диктор видел говорящего на экране, в другом – нет.

Было выяснено, что активнее всего меняются гласные нижнего подъема (в большей степени динамична первая форманта), особенно – /æ/. Более того, было замечено влияние социолингвистических факторов, а именно – расы и пола. В тех случаях, когда фотография того, кому принадлежит голос в наушниках, появлялась на экране, мужчины сильнее проявляли адаптацию (для женщин же ситуация противоположная). Тогда, когда фотографии на экране не было, различий в степени адаптации обнаружено не было (вне зависимости от пола диктора). Что же касается расы, стоит отметить, что у темнокожего человека, голос которого слушали и повторяли дикторы, усредненное значение для а заднего ряда значительно выше, чем у белого. Согласно исследованиям, дикторы в меньшей степени проявляли адаптацию к этому звуку, реализованному темнокожим говорящим (все же более сильно, чем к другим гласным). Однако относительно /æ/ никакой зависимости адаптации от расы выявлено не было.

В целом исследовательница подводит итог: адаптация по формантам не имеет достаточно ощутимой зависимости от расы того, к чьему голосу адаптируется диктор.

1.5. Основные выводы по Главе 1

1. Существует большое количество подходов к трактовке термина коммуникативной адаптации. К тому же, в зависимости от подхода конкретного исследователя или группы ученых различается оценка не только степени адаптации, но и ее направление.
2. Коммуникативная адаптация проявляется на всех уровнях языка, от высших к низшим. Более того, обращается внимание на то, что между процессами производства речи и ее восприятием существует равновесие. Восприимчивость к коммуникативной ситуации в общем смысле оказывает влияние на процесс подстройки речи собеседников, которая, к тому же, происходит вне зависимости от их желания и осознания этого факта
3. О возросшем в последние годы интересе к тематике подстройки акустических характеристик в процессе коммуникации можно судить по количеству работ, проведенных на материале различных языков. Однако, в русской научной литературе можно найти не так много сведений, касающихся исследований в данной области
4. В связи с тем, что исследования в области коммуникативной адаптации набирают популярность, а работ на русском языке, посвященных этой теме, достаточно мало, появилась необходимость создания нового речевого корпуса. Для решения поставленных в рамках этой работы задач были использованы материалы такого корпуса, структура которого позволяла изучить влияние различных факторов на подстройку дикторов в процессе разговора. В качестве критерия оценки факта наличия или отсутствия коммуникативной адаптации были выбраны такие акустические характеристики как формантные частоты ударных гласных.
5. Существует достаточно много методик спектрального анализа акустических характеристик речевого сигнала, как автоматизированных,

так и нет. Метод, используемый в рамках данной работы – построение формантных картин – можно назвать частично автоматизированным.

Глава 2. Исследование формантных характеристик речевой адаптации на материале диалогической речи в русском языке

2.1. Материал исследования – корпус диалогической речи SibLing

В корпус диалогической речи SibLing входят записи диалогов 10 пар однополых братьев и сестер с собеседниками, с которыми они находятся в разной степени знакомства [41]. Возраст основных дикторов – 5 пар мужчин и 5 пар женщин – варьируется в пределах от 23 до 40 лет. Две пары братьев и две пары сестёр являются однойцевыми (монозиготными) близнецами. Основной набор дикторов представлен в таблице 2.1..

Таблица 2.1. Основной набор дикторов. В первом столбце отражены кодовые имена дикторов, во втором – пол (F – женский, M – мужской), в третьем – степень родства, в четвертом – возраст каждого из дикторов в том же порядке, что и кодовые имена в первом столбце

Дикторы	Пол	Родство	Возраст
S01, S02	F	близнецы	26, 26
S03, S04	F	сестры	36, 38
S05, S06	M	братья	36, 38
S07, S08	M	братья	31, 36
S09, S10	F	близнецы	38, 38
S11, S12	M	близнецы	24, 24
S13, S14	F	сестры	28, 30
S15, S16	M	братья	33, 30
S17, S18	F	сестры	23, 25
S19, S20	M	близнецы	23, 23

Каждый из дикторов общался со следующими собеседниками: родной брат или сестра, того же пола и приблизительно того же возраста; близкий друг или подруга того же пола и приблизительно того же возраста; незнакомец

или незнакомка того же пола и приблизительно того же возраста; незнакомец или незнакомка другого пола и приблизительно того же возраста; незнакомец или незнакомка того же пола, занимающий (занимающая) более высокую должность на работе.

Все собеседники являлись носителями русского языка, в настоящее время проживающими в Санкт-Петербурге. Ни один из дикторов не сообщил о каких-либо нарушениях речи и/или слуха. Разница в возрасте собеседников во всех парах, кроме последней (диалог с начальником), не превышала 4 лет.

В общей сложности корпус состоит из 90 диалогов. В студии звукозаписи были получены дорожки формата WAV, 24-бит, с частотой дискретизации 44100 Гц, что говорит о том, что сигнал высокого качества. Для записи использовалась индивидуальная гарнитура AKG HSC 271. Также некоторые диалоги параллельно записывались на микрофон Audio-Technica AT 2050, расположенный между дикторами. Во время коммуникации дикторы либо были разделены непрозрачной ширмой, либо находились в разных комнатах. Эти меры были приняты для того, чтобы собеседники не могли видеть карточки или маршруты друг друга.

В процессе звукозаписи дикторы выполняли 2 типа заданий, которые в общей сложности длились от 25 до 60 минут. Первой задачей, поставленной перед собеседниками, была карточная игра, основанная на поиске сходства в двух колодах из десяти карт. Для этого были использованы карты из игры под названием «Имаджинариум» [15], которая в свою очередь очень схожа с игрой «Диксит», разработанной Жаном-Луи Рубира [14]. На каждой из карточек было изображено несколько различных объектов, объединенных сказочной композицией. Один из участников диалога должен был описать верхнюю карту своему собеседнику, в то время как задача другого состояла в том, чтобы, просмотрев свою колоду, найти карту по крайней мере с двумя совпадающими объектами. Колоды распределялись между собеседниками случайным образом, но на картах обязательно были схожие элементы.

Дикторы по очереди описывали свои карточки. Так, их роли – говорящего и слушающего – менялись несколько раз на протяжении диалога. Игра длилась 10-12 минут, после чего дикторы переходили к выполнению второго задания.

Вторая задача собеседников представляла собой построение маршрутов, в процессе чего они должны были указать друг другу дорогу через локации на схематических картах. У каждого диктора был набор из 4 карт: 2 полные и 2 незаполненные. На полной карте уже был отмечен маршрут, в то время как на незаполненной карте его не было. Более того, некоторые локации отличались или были заменены. Пример задания с маршрутами можно увидеть на рис.2.1.. В то время как диктор, у которого была карта с нарисованным маршрутом, объяснял его своему собеседнику, последнему нужно было изобразить путь на своей карте. После того, как маршрут был объяснен, второй диктор должен был повторить маршрут, который он/она нарисовал/нарисовала. Таким образом, описывая свою заполненную карту, собеседники 4 раза в процессе выполнения второго задания сменили роли говорящего и слушающего.





Рис.2.1. Один из наборов парных карточек (для братьев/ сестер) для выполнения задания на построение маршрута. (Ключевые слова здесь подчеркнуты, но не были каким-либо образом выделены для дикторов) [41]

Каждая пара карт содержала 5 ключевых слов, которые были одинаковыми во всех записях. Все ключевые слова содержали в основном гласные и сонанты (за редким исключением – звонкие щелевые). Такой подбор фонемного состава в словах объясняется тем, что по нему впоследствии легче было сравнивать просодические характеристики. Все ключевые слова приведены в таблице 2.2..

Таблица 2.2. Ключевые слова в заданиях на построение маршрута. В первом столбце - номер маршрута, во втором – соответствующие ключевые слова

Номер маршрута	Ключевые слова
1	ювелирный, военный район, мыловарня, авиаузел, мавзолей
2	равнина, малиновая аллея, земля оленей, руины, взморье
3	муравейник, вороний вольтер, винные земли, заозерье, озеро Ямное
4	ламинария, валериана, мальва, лимон, алоэ

Так как каждый из основных дикторов (братьев, сестер, близнецов) принимал участие в 5 записях, каждый набор маршрутов имел разные конфигурации. При этом ключевые слова оставались одинаковыми, менялись только другие локации и сам маршрут.

После выполнения заданий дикторы заполнили 2 анкеты. Первая, заполняемая дикторами лишь один раз, содержала в себе вопросы личного характера: возраст, пол, уровень образования, профессия, место рождения, родной язык, опыт в занятиях фонетикой иностранных языков; города, где диктор провел свое детство, ходил в школу, учился в университете или колледже и города, где диктор проживал год и больше. Вторая анкета заполнялась после каждого записанного диалога и содержала ряд открытых вопросов, направленных на получение информации о том, чувствовал ли себя диктор комфортно во время этой записи и успешно ли он справился с заданием.

Разметка корпуса включает в себя 7 уровней: реплики, межпаузальные вставки, метки смены ролей, метки периодов основного тона, орфографическая расшифровка и просодическая аннотация, ключевые слова. Также в рамках этой работы в определенных диалогах проведена сегментация на звуки в рамках ключевых слов.

Метки реплик были расставлены автоматически с опорой на амплитуду речевого сигнала. Ошибки программ, сделанные при разметке, были исправлены вручную. Метки смены ролей также были расставлены автоматически на основе межпаузальных вставок.

Орфографическая расшифровка для всех диалогов была составлена лингвистами вручную в соответствии с правилами, ранее разработанными исследователями на кафедре фонетики [42]. В орфографической расшифровке не были использованы знаки препинания, за исключением вопросительного знака, апострофа и дефиса. В аннотации есть указания на

ударения в словах в соответствии со стандартными правилами произношения. Также были отмечены различные нарушения речи, (включая фальстарты, самокоррекцию, оборванные слова, заполненные паузы хезитации), неречевые явления (смех, значимые для дальнейшего анализа речевой адаптации щелчки, кашель). Нестандартные (случайные) произнесения слов (к примеру, «жижда» вместо «жижа») были выделены в тексте расшифровки и к тому же добавлены в специальную таблицу с оговорками в такой форме, чтобы можно было понять, что планировалось и что на самом деле было произнесено.

Далее к орфографической расшифровке была добавлена просодическая аннотация. Используя специальный набор символов, аннотаторы размечали фразовые границы, ядро и тип мелодического движения внутри ядра. Схема, по которой производилась просодическая аннотация, подробно описана в работе Н.Б.Вольской и Т.В.Качковской [59]. Так как ручная просодическая разметка занимает достаточно много времени, на данный момент она была составлена только для фрагментов диалогов, в которых содержится выполнение задания с карточками (примерно 12 минут речи каждого диктора)

Также была составлена фонемная транскрипция в соответствии с правилами нормативного русского произношения [1]. В ее основе лежали орфографические расшифровки, на материале которых был обучен транскриптор, разработанный ранее на кафедре фонетики СПбГУ [28], [7].

2.2. Методика подбора материала из диалогов для исследования формантной адаптации в диалогах

Как было упомянуто выше, материалом для проведения исследования послужил корпус диалогической речи, структура которого подразумевала выполнение дикторами 2 типа заданий. Они были выстроены таким образом, чтобы и в первой, и во второй части диалога собеседники произнесли ключевые слова или сходные фонетические последовательности.

В теории, акустические характеристики звуков внутри этих слов могут меняться в процессе коммуникации.

Для расчета степени адаптации в рамках данной работы использовались ударные гласные. За сходные фонетические последовательности принимаются те слова, в которых правый и левый контекст гласного схожи, но не обязательно совпадают. Рассмотрим слова «пúговица» /púgav'ica/ и «бúковку» /búkafku/. Левый контекст ударного гласного /u/, звуки /b/ и /p/, не совпадают по признаку глухости/звонкости, но обладают одним и тем же местом образования – это губно-губные смычные. Что касается правого контекста, звуков /g/ и /k/, то про них можно сказать то же самое, только по месту образования это будут уже заднеязычные смычные. Для дальнейшего анализа гласного, находящегося в таком окружении, совпадение звуков по месту образования и по типу создаваемой преграды будет важнее различия в признаке звонкости/глухости, так как именно от этого будет зависеть смещение формант в переходных участках гласного [2].

В первой части диалога дикторам было предложено найти карточки со схожими элементами на них и повторить те элементы, которые оказались общими. Для анализа были выбраны те слова, которые повторялись у каждого из дикторов, а также сходные фонетические последовательности. Вторая часть диалога предполагала построение маршрута. В этом отрезке

диалога уже производился поиск заранее определенных ключевых слов (см.раздел 2.1).

Далее обнаруженные у обоих дикторов похожие/одинаковые слова (сходные фонетические последовательности/ключевые слова) отмечались в орфографической расшифровке. После этого вручную проводилась сегментация звукового сигнала: на одном уровне отмечались слова, на другом – звуки (рис.2.2.).

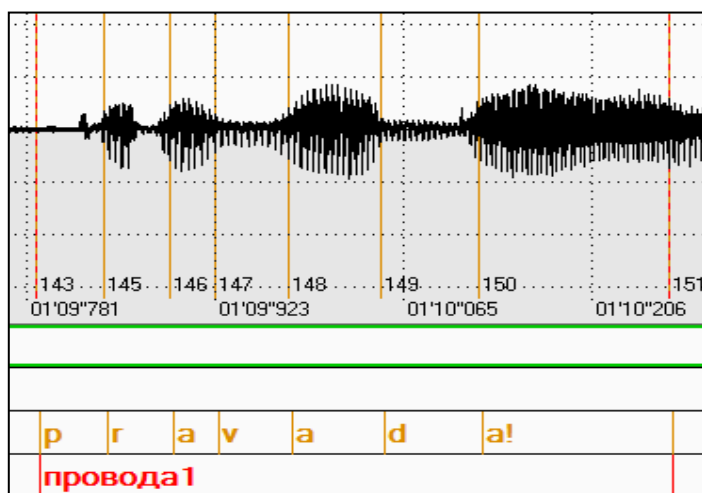


Рис.2.2. Сегментация звукового сигнала на 2 уровнях – словесном и звуковом – на примере слова «провода». Метки красного цвета обозначают расположение слова на звуковом сигнале, а метки желтого цвета – границы звуков. Ударение обозначено цифрой «1» на словесном уровне и восклицательным знаком на звуковом.

Метод ручной сегментации позволяет добиться высокого качества полученных впоследствии меток [17]. Также одним из способов обозначения границ сегментов речи (слов, звуков), использованным в рамках данной работы, является сервис WebMAUS [16]. Некоторые диалоги, сегментация которых производилась посредством этого инструмента, потребовали ознакомления с дополнительным программным обеспечением на языке программирования Python, подготавливающим сигнал к обработке, а также с самим инструментом.

Так как и ручная сегментация, и работа с сервисом Webmaus требуют большого количества времени на обработку сигнала, в общей сложности в рамках данной работы было обработано 7 диалогов из корпуса SibLing, данные о которых представлены в таблице 2.3. Длительность всего обработанного материала составляет 4 часа 16 минут. Дикторы были подобраны таким образом, чтобы по результатам акустического анализа речевого сигнала можно было судить о зависимости коммуникативной адаптации от степени знакомства собеседников.

Для дальнейшего проведения акустического анализа нужны были отдельные сегменты, а именно гласные в ударных слогах. Для обработки были написаны скрипты на языке Python которые вырезали звуки для проведения формантного анализа, использующие ранее составленную сегментацию на звуковом и словесном уровнях.

Все звуки, полученные на выходе были дополнительно приведены к единой частоте дискретизации 22 050 Гц и к максимальной амплитуде. Такая нормализация речевого сигнала была проведена для того, чтобы избежать сбоев в работе программного обеспечения, используемого для дальнейшего формантного анализа.

Таблица 2.3. Данные об обработанных диалогах и собеседниках, принимавших в них участие. В первом столбце находится номер диалога согласно присвоенной ему кодировке, во втором и пятом столбцах – дикторы, имена которых также закодированы, в третьем и шестом – пол дикторов, в четвертом и седьмом – возраст, восьмой столбец содержит информацию о степени знакомства информантов.

Номер диалога	Собеседник №1	Пол	Возраст	Собеседник №2	Пол	Возраст	Степень знакомства
D11	S03	Ж	36	S04	Ж	38	сестры
D12				Z11	Ж	36	друзья
D13				Z12	Ж	35-40	незна-комцы
D14				Z13	М	38	незна-комцы разного пола
D15				Z14	Ж	66	незна-комцы разного социального статуса (Z14 – началь- ник)
D51	S11	М	24	S12	М	24	близне-цы
D81	S17	Ж	23	S18	Ж	25	сестры

2.3. Расчет формантных характеристик

Для обоснования дальнейших действий, производимых с речевым сигналом, необходимо обратиться к акустической теории речеобразования Г. Фанта [19]. Речевой аппарат условно можно поделить на 2 отдела в соответствии с функциями, за которые они отвечают. В модели Г.Фанта речевой тракт человека представлен в качестве схемы «источник-фильтр» [13].

Если рассматривать конкретно гласные звуки через призму этой теории, то можно говорить о том, что существует источник звука – глоттальная волна или голос – на которую накладываются модуляции. Проходя через «трубу», образуемую между голосовой щелью и губами, звук претерпевает изменения, которые зависят от расположения артикуляторных органов. На выходе получается сигнал, в котором усилены определенные частоты за счет умножения на значения передаточной функции (рис. 2.3).

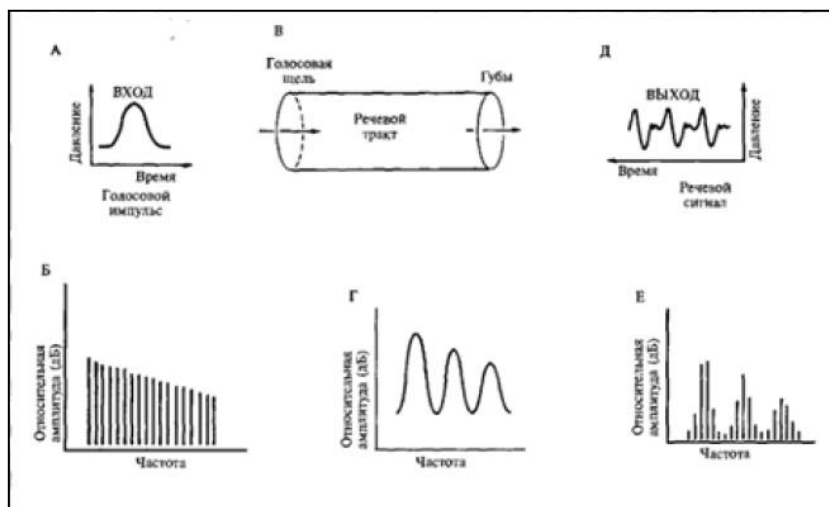


Рис.2.3. Изменение речевого сигнала в процессе производства гласных [13].

Формантами называются резонансные максимумы передаточной функции, иначе – собственные частоты резонаторов, наиболее усиленные в спектральной картине звука. Эти частотные максимумы отвечают за качество звука.

Опираясь на акустическую теорию речеобразования [19] и на теорию речевой адаптации [34], можно сказать о том, что подстройка собеседников под речевое поведение друг друга может оказать влияние на изменение формант. Это произойдет потому, что дикторы начнут смещать артикуляцию гласных с привычных для их речи на те, которые акустически близки к местам артикуляции собеседника.

Важно упомянуть, что формантные характеристики нельзя назвать строго определенными в пространстве точками. Это скорее набор точек, сконцентрированных в определенной зоне речевого тракта. Поэтому в спектре можно обнаружить целый набор чисел, соответствующий формантным частотам.

В рамках данной работы приведённые к единому формату звуки обрабатывались программой, позволяющей рассчитывать спектральные характеристики звуков [4], [5], [6].

Результат работы программы выводился в формате книг Excel.

A	B	C	D	E	F
37	я1мное	a!.wav	672	1964	2421
37	я1мное	a!.wav	684	1840	2393
55	я1мное	a!.wav	483	1747	2149
55	я1мное	a!.wav	577	1644	2103
37	я1мное	a!.wav	726	1636	2313
55	я1мное	a!.wav	662	1525	2118
37	я1мное	a!.wav	768	1513	2317
55	я1мное	a!.wav	775	1455	2148
3	мы2лова1	a!.wav	702	1421	2064
37	я1мное	a!.wav	796	1366	2313
37	я1мное	a!.wav	787	1257	2313

Рис.2.4. Отображение данных спектрального анализа в таблице Excel. В столбце А содержатся номера вырезанных звуков, в столбце В – слова, из которых эти звуки взяты, в столбце С указан проанализированный ударный звук, в столбцах D, E и F находятся значения для первой (F1), второй (F2) и третьей (F3) формант соответственно.

Как видно из рис.2.4, для каждого звука программой выводились спектральные характеристики последовательно на протяжении гласного. Следующим этапом являлось усреднение формантных значений в Excel для дальнейшего построения формантных картин.

В дальнейшем проводился зрительный анализ получившихся графиков. Также при оценке характера изменения положения звуков на формантных картинах использовалось Евклидово расстояние. Это метрика, которая позволяет провести расчет, на сколько в Евклидовом пространстве 2 точки сблизилась или отделились [9].

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Рис.2.5. Формула для расчета Евклидова расстояния «между точками x и y в n-мерном пространстве» [9].

В подсчетах в рамках этой работы положительное значение сигнализировало об уподоблении акустических характеристик двух дикторов, а отрицательное значение – об обратном ему процессе, дивергенции.

2.4. Результаты сравнения формантных характеристик

Для каждой пары дикторов были построены формантные картины, соответствующие положению формантных частот в начале и в конце записи. На каждом из графиков на оси X находятся значения для F2, а на оси Y – значения для F1. Такой вид графиков является удобным для дальнейшего рассмотрения формантных картин, так как его можно наложить на сагиттальный разрез. Это действие позволяет оценить как изменилась артикуляция отдельно взятого звука и каким он стал по качеству, то есть что произошло с его рядом и подъемом.

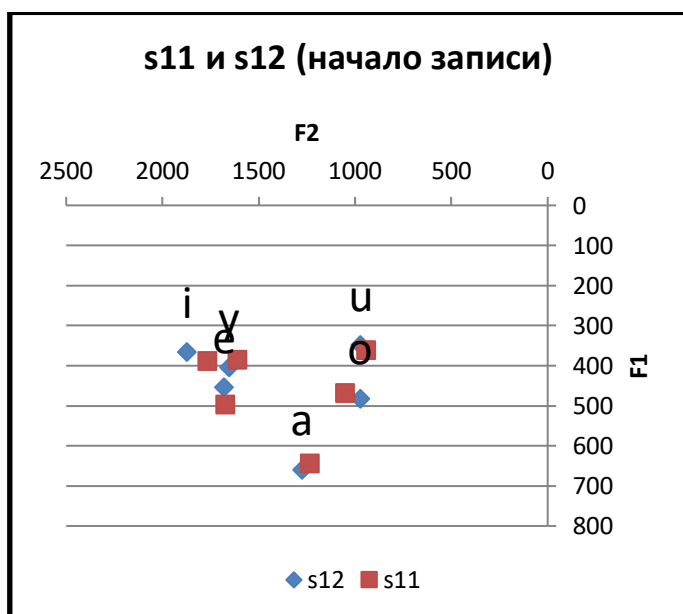


Рис.2.6. Формантная картина гласных дикторов s11 и s12 в начале записи

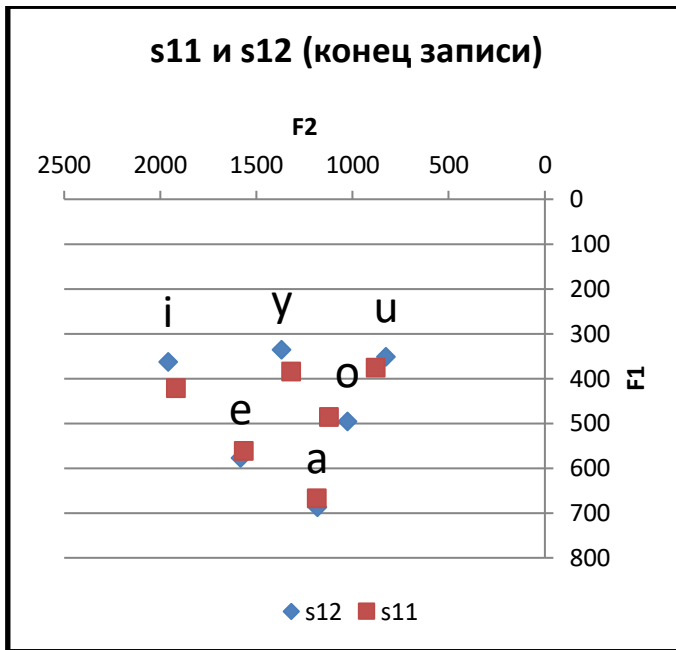


Рис. 2.7. Формантная картина гласных дикторов s11 и s12 в конце записи

На двух данных графиках, отражающих формантные картины гласных пары близнецов мужского пола (s11 и s12) в начале и в конце диалога, можно отметить смещение формантных частот как у первого диктора, так и у второго. Более того, изменение произошло не в сторону частот одного из дикторов. Можно отметить взаимное выравнивание формантных картин.

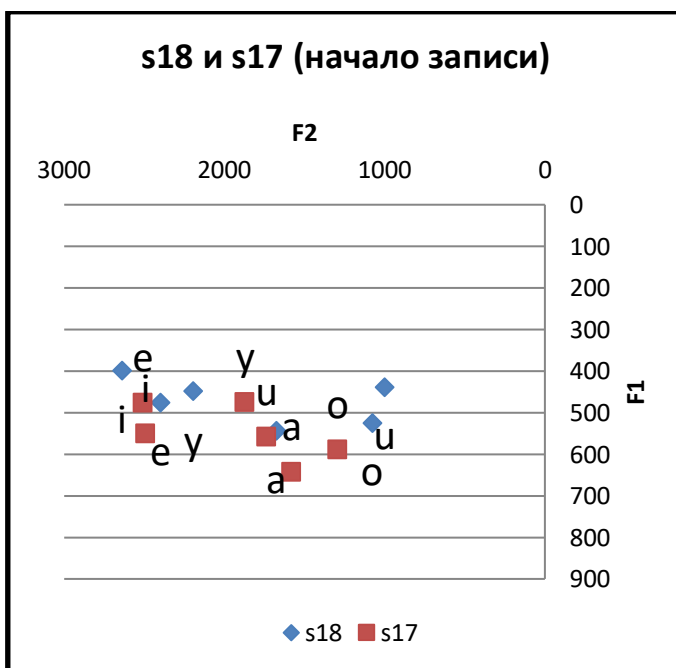


Рис.2.8. Формантная картина гласных дикторов s17 и s18 в начале записи

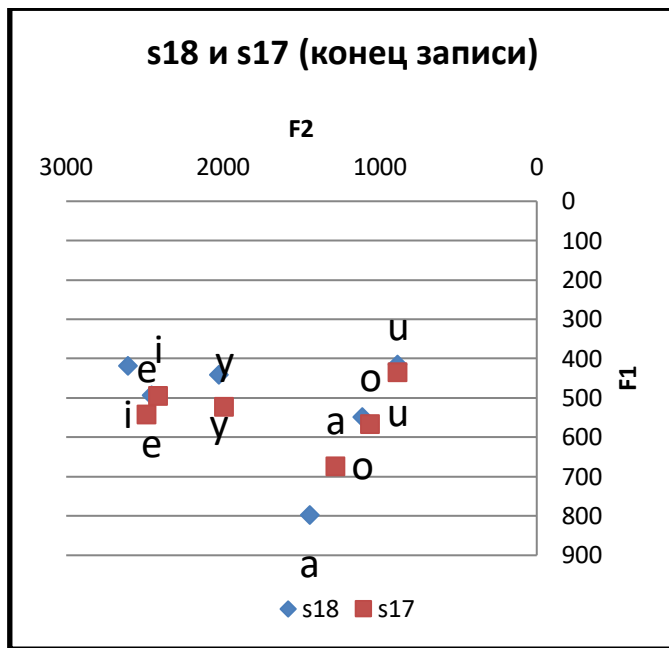


Рис.2.9. Формантная картина гласных дикторов s17 и s18 в конце записи

На примере диалога двух сестер (s17 и s18) наблюдается выравнивание формантных картин гласных. График, соответствующий положению формант дикторов в конце записи, является более упорядоченным. В целом все звуки рассредоточились по условному «треугольнику гласных». Более того, точки, соответствующие гласным заднего ряда, встали более близко друг к другу. Также наблюдается сближение формантных значений гласных переднего ряда, но уже не такое значительное. Однако, об адаптации по гласному /a/ говорить нельзя.

Далее был проведен анализ серии диалогов одного из дикторов с собеседниками с разной степенью знакомства. В этих диалогах одна из сестер (s03) вступала в беседу с другой сестрой, подружкой, незнакомкой, незнакомцем, а также с человеком более высокого социального статуса.

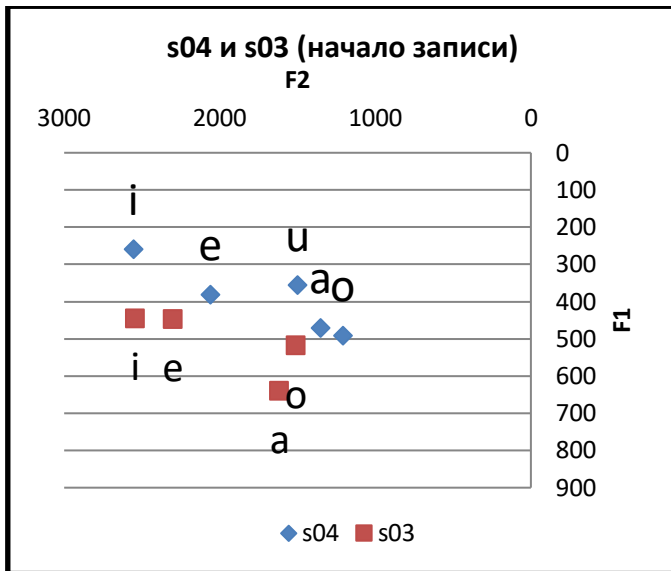


Рис.2.10. Формантная картина гласных дикторов s03 и s04 в начале записи

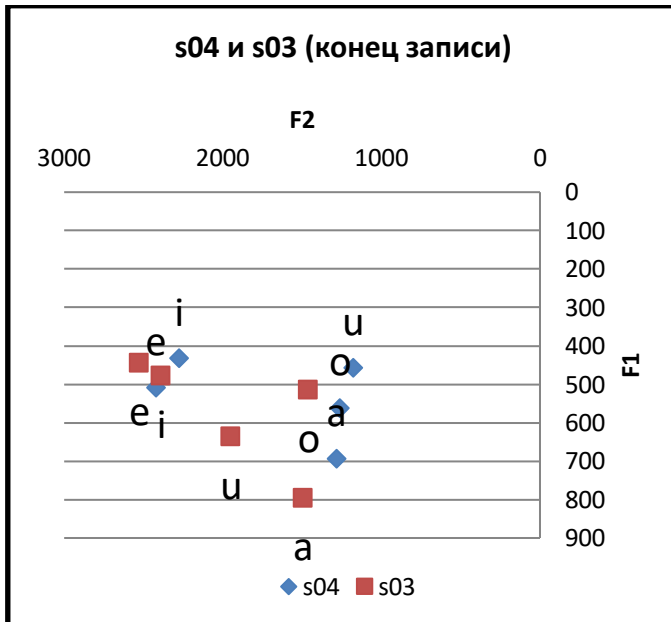


Рис.2.11. Формантная картина гласных дикторов s03 и s04 в конце записи

Приведенные выше графики отражают взаимное расположение гласных двух сестер (s03 и s04) в начале и в конце диалога. Все гласные сблизилась в процессе коммуникации.

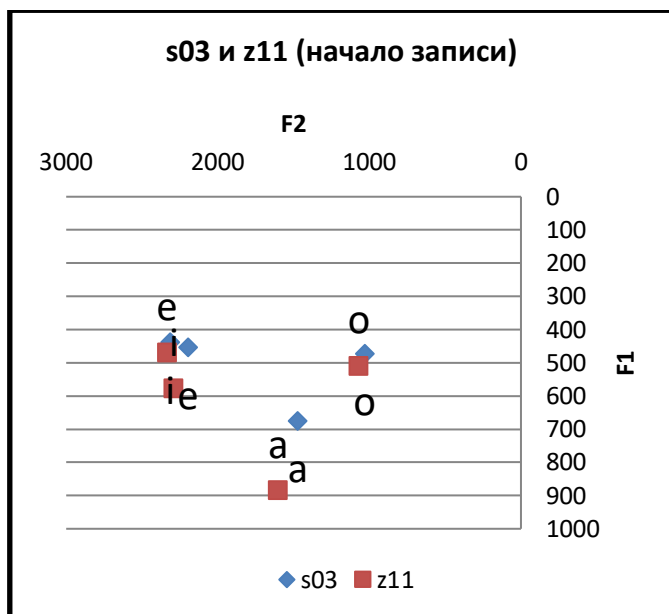


Рис.2.12. Формантная картина гласных дикторов s03 и z11 в начале записи

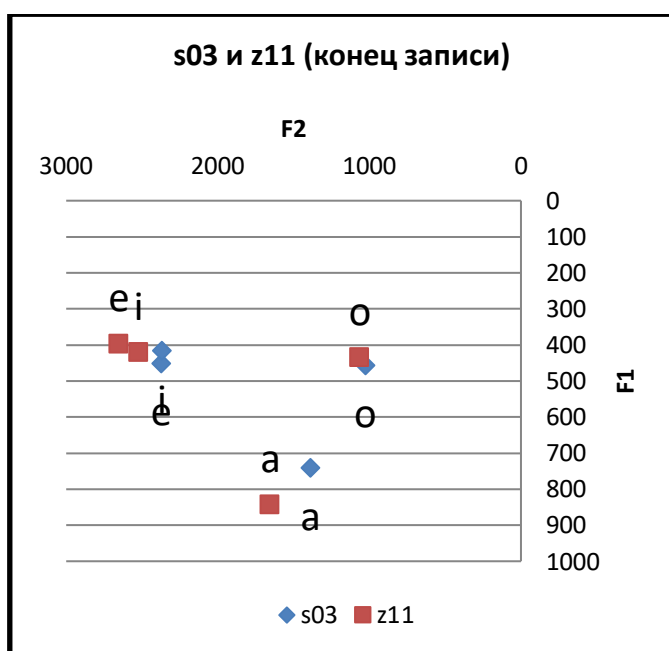


Рис.2.13. Формантная картина гласных дикторов s03 и z11 в конце записи

Диалог одной из сестер (s03) с подругой (z11) представляет собой интересный пример влияния степени знакомства собеседников на скорость коммуникативной адаптации. Формантные картины и в начале записи, и в конце, достаточно схожи. Из этого можно сделать вывод о том, что адаптация уже произошла в первой части диалога.

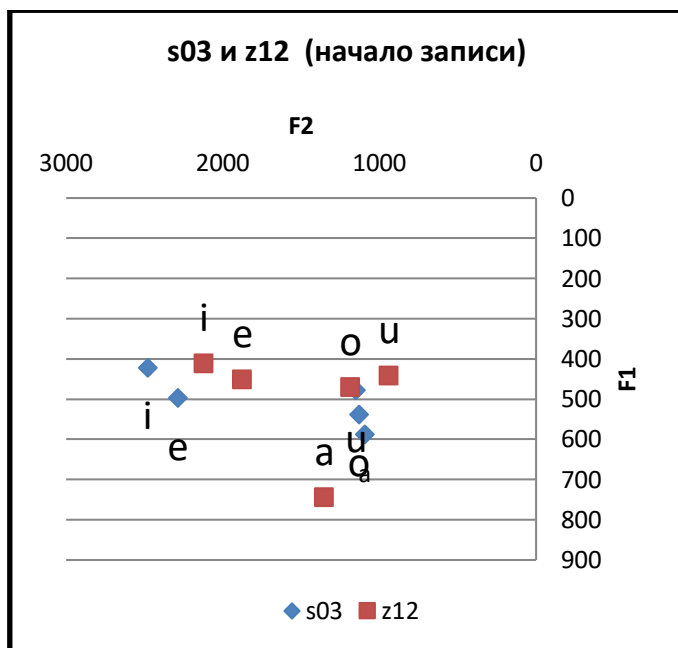


Рис.2.14. Формантная картина гласных дикторов s03 и z12 в начале записи

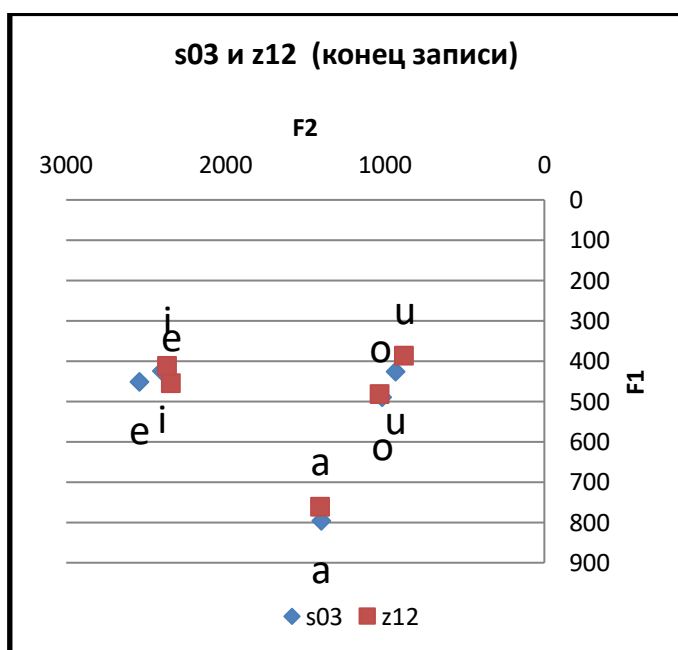


Рис.2.15. Формантная картина гласных дикторов s03 и z12 в конце записи

Анализ формантных картин показывает, что в диалоге между сестрой (s03) и незнакомкой (z12) также произошла коммуникативная адаптация. Особенно приблизились друг к другу в процессе диалога точки, соответствующие гласным заднего ряда, а также гласному нижнего подъема - /a/.

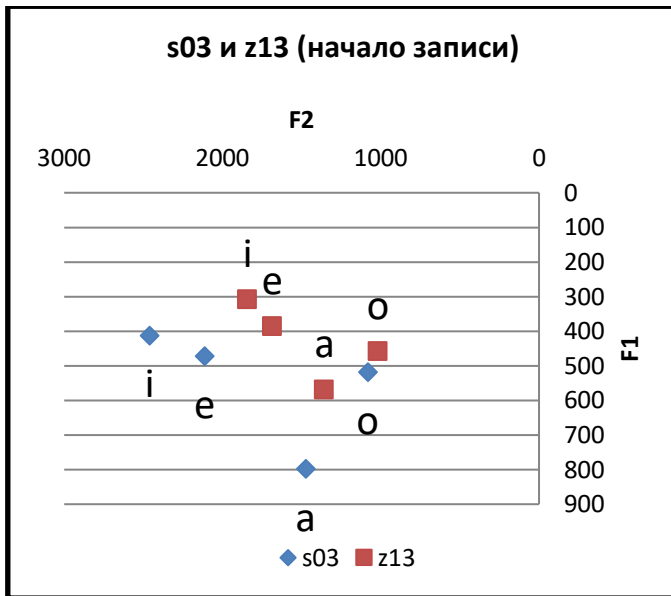


Рис.2.16. Формантная картина гласных дикторов s03 и z13 в начале записи

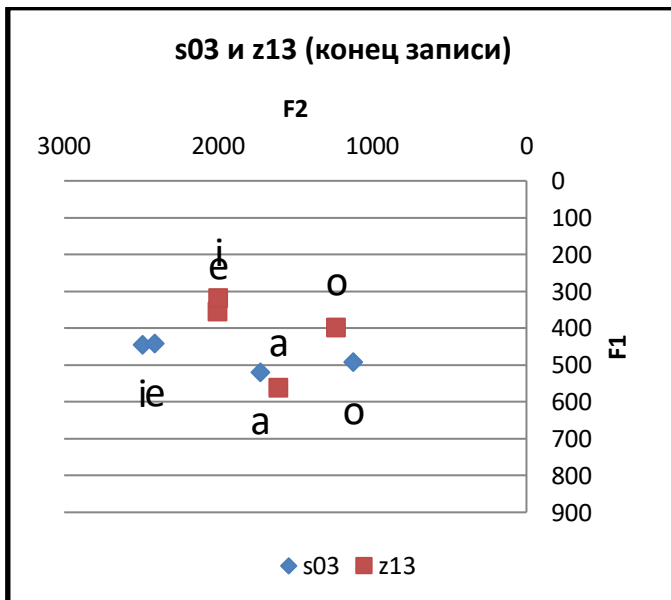


Рис.2.17. Формантная картина гласных дикторов s03 и z13 в конце записи

Так как диктор-незнакомец (z13) мужского пола и повышение ЧОТ, и формант гласных соответственно для него, скорее всего, нелегкая задача, сестра (s03) в процессе диалога подстраивается под его формантные значения. При этом она достаточно сильно сдвигает свой гласный /a/ в диапазон частот, доступный диктору-незнакомцу. Более того, незнакомец смещает фокус артикуляции гласных /e/ и /i/, меняя их качество на более переднее.

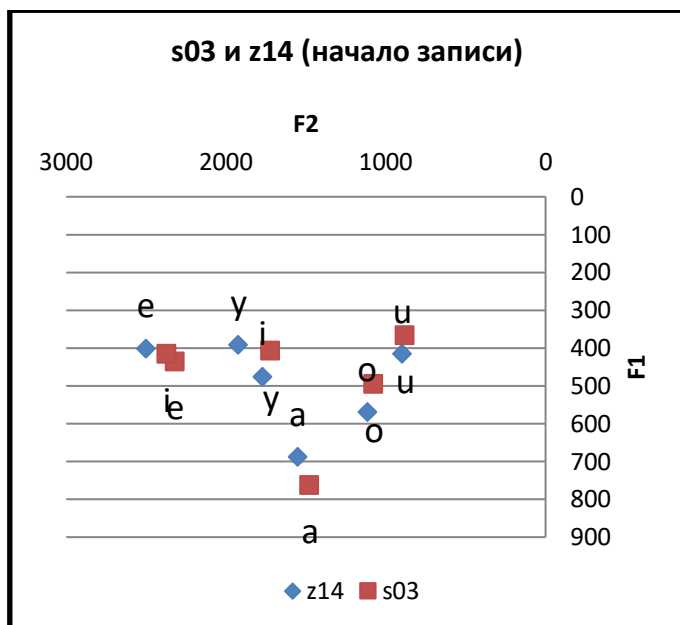


Рис.2.18. Формантная картина гласных дикторов s03 и z14 в начале записи

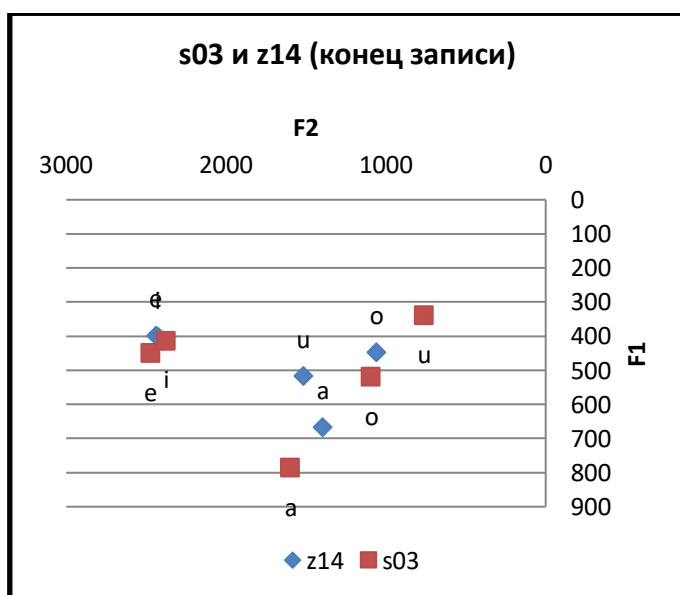


Рис.2.19. Формантная картина гласных дикторов s03 и z13 в конце записи

Что касается разговора сестры (s03) с человеком более высокого социального статуса (начальником, z14), скорее можно сделать выводы о том, что произошел процесс, обратный коммуникативной адаптации. Опираясь и на формантные картины, можно говорить о том, что почти все гласные дикторов отдалились друг от друга. Однако, гласные переднего ряда /e/ и /i/ (последний – в большей степени) сблизилась в процессе коммуникации.

Для более точного описания происходивших во время диалогов изменений спектральных характеристик гласных использовалось Евклидово расстояние (таблица 2.4.).

Таблица 2.4. Данные о Евклидовом расстоянии по всем проанализированным диалогам. Номер диалога и данные о степени знакомства собеседников представлены в 1 столбце, проанализированные фонемы – во втором, значения Евклидова расстояния в начале и в конце диалога – в третьем и четвертом столбцах соответственно, разница между значениями из третьего и четвертого столбиков отражена в пятом

Диалог	Звук	Евклидово расстояние в начале диалога	Евклидово расстояние в конце диалога	Разница в Евклидовом расстоянии
D11 (диалог двух сестер)	a	312.90	232.76	80.15
	e	247.86	122.14	125.72
	o	302.55	201.61	100.95
	i	186.74	120.86	65.87
D12 (диалог между одной из сестер и подругой)	a	246.79	284.96	-38.17
	e	139.23	284.08	-144.85
	o	53.05	41.18	11.87
	i	141.06	154.08	-13.02
D13 (диалог между одной из сестер и незнакомкой)	a	301.35	34.63	266.72
	e	415.44	202.05	213.40
	o	149.46	13.46	136.00
	i	359.61	35.11	324.50
D14 (диалог между одной из сестер и незнакомцем)	a	256.53	126.83	129.83
	e	434.69	501.86	-67.17
	o	87.08	142.96	-55.88
	i	627.10	432.14	194.96

D15 (диалог между одной из сестер и начальницей)	a	107.48	233.67	-126.18
	e	188.28	62.34	125.94
	o	84.72	81.08	3.64
	u	50.23	778.51	-728.27
	i	447.75	54.83	392.93
D51 (диалог пары близнецов мужского пола)	a	54.88	44.20	10.68
	e	125.15	46.17	78.98
	o	72.17	76.71	-4.55
	u	63.78	35.88	27.90
	y	292.00	50.92	241.08
	i	155.32	111.88	43.44
D81 (диалог двух сестер)	a	140.73	208.32	-67.59
	e	109.02	57.12	51.90
	o	224.76	54.25	170.50
	u	745.23	20.60	724.63
	y	325.66	90.01	235.65
	i	211.45	210.02	1.43

Приведенные в таблице данные подтверждают и дополняют выводы, сделанные ранее на основании визуального анализа формантных картин.

Так, наибольшие изменения во взаимоположении гласных в начале и в конце диалога между s11 и s12 претерпели гласные верхнего подъёма во всех трех рядах, /i/, /y/ и /u/, а также гласный /e/ среднего подъёма переднего ряда.

Результаты анализа записи диалога двух сестер, s17 и s18, показывают, что сильнее всех поменялось расстояние между гласными заднего ряда /o/ и /u/, что можно сказать и про /y/.

В процессе коммуникации следующей пары сестер, s03 и s04, больше всего сблизилась, опять же, гласные заднего ряда /o/, а также достаточно сильно сократилось расстояние между гласными переднего ряда /e/.

Евклидово расстояние между гласными, взятыми из диалога сестры, s03, и ее подруги, z11, в целом отрицательное. Это может говорить о том, что формантные частоты девушек сблизилась уже в первой части диалога, а затем лишь незначительно смещались, в целом оставаясь схожими. Можно заметить, что Евклидово расстояние в рамках этого диалога колеблется в достаточно маленьком диапазоне. Исключение составляет только случай гласного /e/.

Говоря о диалоге сестры, s03, и незнакомки, z12, можно отметить достаточно высокую степень коммуникативной адаптации, которая затронула все гласные без исключения. Особенно сильно сблизилась характеристики гласных /a/, /i/ и /e/.

Что касается диалога сестры, s03, и незнакомца, z13, снова стоит отметить сближение гласных /a/ и /i/. Гласные же среднего подъема, /e/ и /o/, отдалились в процессе коммуникации, но незначительно (значения колеблются примерно в том же диапазоне, что и в диалоге сестры и подруги).

О диалоге сестры, s03 с незнакомкой с более высоким статусом (начальницей), z14, можно в целом говорить как о примере дивергенции. Однако гласные переднего ряда, /i/ и /e/ достаточно сблизилась в процессе диалога. Изменения в положении гласных /o/ можно считать незначительными, так как Евклидово расстояние между ними, хоть и положительное, но является достаточно малым.

2.5. Основные результаты и выводы

Итак, опираясь на расчет Евклидова расстояния и анализ формантных картин гласных, можно сделать следующие выводы:

1) В большей части случаев наблюдается взаимное смещение формантных характеристик гласных, результатом которого является перестройка треугольников гласных у обоих дикторов. Соответственно, наблюдается взаимная подстройка собеседников в процессе диалога.

К примеру, в случае диалога одной из сестер с незнакомцем, можно было наблюдать интересное явление: сестра переместила фокус артикуляции гласного /a/ в диапазон значений, наиболее близкий для второго диктора. В то же время незнакомец, оставляя неизменным качество гласного /a/, стал реализовывать гласный /i/ как более передний.

2) По предварительным данным, степень знакомства способна влиять на скорость коммуникативной адаптации (это можно наблюдать на примере диалога сестры с подругой);

Более того, наблюдается связь между степенью знакомства и степенью адаптации. В диалогах близких родственников (братьев и сестер) наблюдается или изначальное сходство формантных характеристик, или их сближение в процессе коммуникации. При общении с незнакомцами скорее можно говорить о конвергенции, нежели о дивергенции.

Также в диалоге одной из сестер был замечен случай дивергенции. Это явление можно связать с тем, что собеседники были разного возраста. К тому же на степень адаптации могла повлиять разница в социальном статусе.

3) Степень подстройки акустических характеристик в некоторой степени зависит от качества гласного. Основываясь на зрительном анализе

формантных картин, можно говорить о том, что в большей степени дикторы адаптируются друг к другу по огубленным гласным заднего ряда (/o/, /u/), а также активно меняют, подстраиваясь друг под друга, местоположение гласных /a/ и /i/. Также стоит отметить и то, что зачастую дикторы смещают произношение сразу всех кардинальных гласных. Более того, встречались и случаи, смещения гласных в пределах целого подъема: к примеру, к адаптации по /i/ и /u/ присоединялся гласный /y/; вместе с гласным /e/ менял свои характеристики и гласный /o/. То же самое можно сказать и про смещение фокуса артикуляции гласных в пределах ряда: к изменению качества /i/ зачастую присоединяется динамика в качестве /e/.

Опираясь на данные о Евклидовом расстоянии (рис.19), можно сказать, что меньше всего отрицательных значений, то есть отдаления фокусов артикуляции гласных в процессе коммуникации, наблюдается у гласного /i/, на втором месте находится гласный /o/, /a/ и /e/ занимают третье и четвертое места соответственно. Аналогичный рейтинг можно составить и анализируя максимальные значения Евклидова расстояния: наиболее сильно в процессе диалога у собеседников сближаются гласные /i/, чуть меньше - /a/, еще меньше - /e/, а на последнем месте находится гласный /o/.

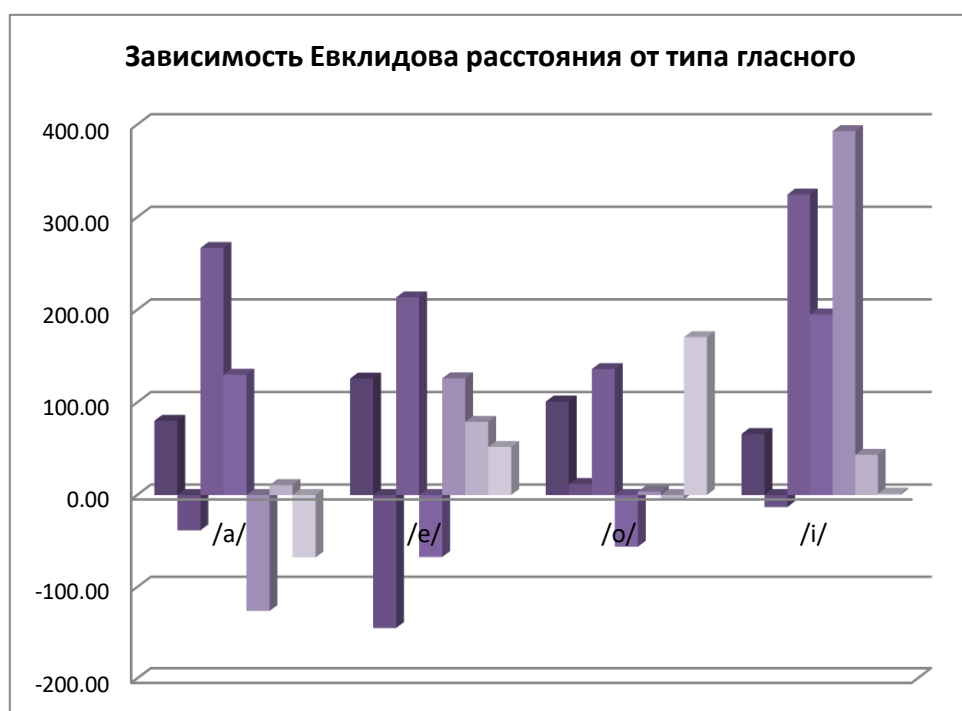


Рис.2.20. Зависимость степени адаптации от качества гласного во всех диалогах. По оси у отображены значения, полученные после расчета разницы Евклидова расстояния. Столбцы, отражающие эти значения в каждом отдельно взятом диалоге, для наглядности объединены в 4 группы для 4 разных фонем

4) На данном этапе исследования недостаточно данных, по которым можно было бы уверенно говорить о зависимости коммуникативной адаптации от пола собеседника. Предварительно можно сказать о том, что дикторы активнее меняют акустические характеристики, подстраиваясь под собеседника того же пола. Однако, нужно учитывать, что возможно влияние целой совокупности факторов (пол, возраст, социальный статус), вследствие чего сложно будет говорить об адаптации даже в случае, когда собеседники одного пола.

Результаты исследования были представлены на XXIII Открытой конференции студентов-филологов в СПбГУ, сборник по итогам которой на данный момент находится в печати [10].

Заключение

В данной исследовательской работе внимание было сосредоточено на динамике акустических характеристик под влиянием коммуникативной адаптации в диалоге. Проводился анализ данных, касающихся спектральных характеристик ударных гласных в сходных фонетических последовательностях.

Материалом послужил корпус диалогической речи SibLing, разработанный на кафедре фонетики СПбГУ. Все необходимые сегменты речи (ударные звуки, слова), были отсегментированы вручную, что позволило повысить уровень качества разметки.

Проводилось усреднение значений по первой и второй формантам ударных гласных из таблиц Excel, полученных после автоматизированного спектрального анализа. Для того, чтобы проводить дальнейшее изучение динамики акустических характеристик от первой части диалога ко второй, был выбран метод построения формантных картин. Подтверждение результатов зрительного анализа проводилось путем расчета Евклидова расстояния между гласными.

По итогам сравнения формантных картин можно сделать вывод о взаимном характере подстройки акустических характеристик. В большей части случаев перестраивались треугольники гласных обоих дикторов.

Также по предварительным данным можно говорить о влиянии степени знакомства как на степень, так и на скорость адаптации. Подстройка акустических характеристик является наиболее быстрой и сильной в случаях, когда между собеседниками наблюдается наименьшее расстояния с точки зрения социолингвистических характеристик (уровень близости межличностных отношений, возраст, социальный статус).

Было обнаружено, что степень адаптации зависит также и от фонемного статуса гласного. Так, в наибольшей степени в речи собеседников подстраиваются огубленные гласные заднего ряда /o/, /u/, а также передний гласный /i/ и гласный нижнего подъема /a/.

Что касается пола собеседников, то можно говорить о том, что влияние этого социолингвистического фактора на степень коммуникативной адаптации тоже может рассматриваться. Однако, на данном этапе исследования для того, чтобы сделать надежные выводы, недостаточно данных. По предварительным данным можно говорить о том, что наиболее активная подстройка акустических характеристик наблюдается в диалогах собеседников одного и того же пола. Однако, стоит учесть возможную совокупность влияния различных социолингвистических факторов – не только пола, но и возраста, социального статуса, степени знакомства.

Библиографический список

1. Аванесов Р. И. Русское литературное произношение. М.: Просвещение, 1984. 383 с.
2. Бондарко Л. В. Фонетика современного русского языка. СПб.: Изд-во С.-Петербургского Университета, 1998. 276 с.
3. Гураков И.А., Костюченко Е.Ю., Новохрестова Д.И., Силич М.П. Алгоритм выделения формант и поиска выровненных фрагментов при подготовке к проведению фоноскопической экспертизы. // Доклады ТУСУРа, 2018. Т. 21. Вып.2. с.48-53
4. Евдокимова В. В. Вариативность формантной картины гласных в разных видах речи // Труды второго междисциплинарного семинара "Анализ разговорной русской речи"(АРЗ-2008). Санкт-Петербургский государственный университет аэрокосмического приборостроения, 2008. С. 49-54.
5. Евдокимова В. В. Сравнительный анализ современных методов построения формантных картин гласных // Вестник Санкт-Петербургского университета. Язык и литература, 2010. Вып. 2. С.80-86
6. Евдокимова В. В., Кочаров Д. А., Скредин П. А. Метод построения формантных картин для исследования фонетических характеристик гласных // Информатика и автоматизация, 2020. Т. 19. Вып. 2. С. 302-329.
7. Евдокимова В. В., Скредин П. А., Чукаева Т. В. Автоматический адаптивный фонетический транскриптор для русского языка // Труды седьмого междисциплинарного семинара «Анализ разговорной русской речи (АРЗ-2017)», 2017. С. 32-39.
8. Евдокимова В.В. – Акустический анализ речевого сигнала: методическое пособие по семинарским и практическим занятиям // СПб.: Издательство СПбГУ, 2014. 104 с.

9. Евклидово расстояние // [Электронный ресурс] / URL: <http://statistica.ru/glossary/general/evklidovo-rasstoyanie/> (дата обращения – 31.05.2021)
10. Зими́на С.В. Акустические характеристики коммуникативной адаптации в диалоге в сходных фонетических последовательностях // Сборник по итогам XXIII Открытой конференции студентов-филологов в СПбГУ. В печати.
11. Зотина Е. В. Коммуникативная подстройка как фактор построения успешной коммуникации (на материале записей образцов голоса и речи) // Ученые записки Казанского юридического института МВД России, 2017. Т. 2. Вып. 4. С.115-121
12. Иванов А.В., Трушин В.А., Маркелова Г.В., Рева И.Л. Исследование спектра формант форсированной речи // Научный вестник НГТУ, 2015. Т.61. Вып.4, 2015. с.63-73
13. Кодзасов С.В., Кривнова О.Ф. Общая фонетика: Учебник. М.: Российский государственный гуманитарный университет, 2001. с. 103-104
14. Настольная игра «Диксит» // [Электронный ресурс] / URL: [https://ru.wikipedia.org/wiki/Dixit_\(%D0%B8%D0%B3%D1%80%D0%B0\)](https://ru.wikipedia.org/wiki/Dixit_(%D0%B8%D0%B3%D1%80%D0%B0)) (дата обращения – 31.05.2021)
15. Настольная игра «Имаджинариум» // [Электронный ресурс] / URL: <https://ru.wikipedia.org/wiki/%D0%98%D0%BC%D0%B0%D0%B4%D0%B6%D0%B8%D0%BD%D0%B0%D1%80%D0%B8%D1%83%D0%BC> (дата обращения – 31.05.2021)
16. Сервис для автоматической разметки WebMAUS // [Электронный ресурс] / URL: <https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/WebMAUSGeneral> (дата обращения: 5.05.2021)

- 17.Скрелин П. А. Сегментация и транскрипция. Автономная некоммерческая организация" Издательство Санкт-Петербургского государственного университета", 1999.
- 18.Сорокин В.Н., Леонов А.С., Макаров И.С. Устойчивость оценок формантных частот // Речевые технологии, 2009. с.3-21
- 19.Фант Г. Акустическая теория речеобразования. М.: Наука, 1964. 284 с.
- 20.Шевченко Т. И., Галочкина И. Е. Навстречу согласию: фонетическая адаптация собеседников в диалоге //Вестник Московского государственного лингвистического университета. Гуманитарные науки, 2016. Вып. 1 (740). С.185-203.
- 21.Babel M. Selective vowel imitation in spontaneous phonetic accommodation // UC Berkeley PhonLab Annual Report, 2009. Vol. 5. N 5. Pp.163-194.
- 22.Bailly G., Lelong A. Speech dominoes and phonetic convergence //11th Annual Conference of the International Speech Communication Association 2010 (Interspeech 2010), 2010. Pp.1153-1156.
- 23.Bilous F. R., Krauss R. M. Dominance and accommodation in the conversational behaviours of same-and mixed-gender dyads // Language & Communication, 1988. Pp.183-194.
- 24.Branigan H. P., Pickering M. J., Cleland A. A. Syntactic co-ordination in dialogue //Cognition, 2000. Vol. 75. N 2. Pp. B13-B25.
- 25.Crystal, D. 1994. Documenting rhythmical change // Jack Windsor Lewis (ed.), Studies in General and English Phonetics, 1994. Pp.174–179.
- 26.Dragojevic M., Gasiorek J., Giles H. Accommodative strategies as core of the theory // Communication accommodation theory: Negotiating personal relationships and social identities across contexts, 2016. Pp. 36-59
- 27.Elhami A. Communication Accommodation Theory: A Brief Review of the Literature //Journal of Advances in Education and Philosophy, 2020. Vol. 4. Pp. 192-200.

28. Evdokimova V., Skrelin P., Chukaeva T. Automatic phonetic transcription for Russian: Speech variability modeling // International Conference on Speech and Computer. – Springer, Cham, 2017. Pp. 192-199.
29. Fant G., Liljencrants J., and Lin Q. A four-parameter model of glottal flow // STL-QPSR, 1985. Vol. 4. Pp. 1–13
30. Gallois C., Giles H. Accommodating mutual influence in intergroup encounters // Progress in communication sciences, 1998. Pp. 135-162.
31. Gallois, C., Ogay, T., & Giles, H. Communication Accommodation Theory: a look back and a look ahead // W. B. Gudykunst (Ed.) Theorizing about intercultural communication, 2005. Pp. 121-148.
32. Gasiorek J., Giles H. Effects of inferred motive on evaluations of nonaccommodative communication // Human Communication Research, 2012. Vol. 38. N 3. Pp. 309-331.
33. Giles H. Accommodation theory: Some new directions // York papers in Linguistics, 1980. Vol.9. N 450. Pp. 105-136.
34. Giles H. et al. Speech accommodation theory: The first decade and beyond // Annals of the International Communication Association, 1987. Vol. 10. N 1. Pp. 13-48.
35. Giles, H., Bourhis, R. Y., & Taylor, D. M. Towards a theory of language in ethnic group relations // Language, ethnicity and intergroup relations, 1977
36. Giles H., Gasiorek J. Intergenerational communication practices // Handbook of the psychology of aging. Academic Press, 2011. Pp. 233-247.
37. Giles H., Powesland P. F. Speech style and social evaluation. Academic Press, 1975.
38. Giles, H., & Ogay, T. Communication Accommodation Theory // B. B. Whaley & W. Samter (Eds.), Explaining communication: Contemporary theories and exemplars, 2007. Pp. 293–310.
39. Godfrey J. J., Holliman E. C., McDaniel J. SWITCHBOARD: Telephone speech corpus for research and development // Acoustics, Speech, and Signal

- Processing, IEEE International Conference on, IEEE Computer Society, 1992. Vol. 1. Pp. 517-520.
40. Hewett D. G., Watson B. M., Gallois C. Communication between hospital doctors: Underaccommodation and interpretability // *Language & Communication*, 2015. Vol. 41. Pp. 71-83.
 41. Kachkovskaia T. et al. SibLing corpus of Russian dialogue speech designed for research on speech entrainment // *Proceedings of The 12th Language Resources and Evaluation Conference*, 2020. Pp. 6556-6561.
 42. Kachkovskaia, T., Kocharov, D., Skrelin, P., and Nina, V. CoRuSS—a new prosodically annotated corpus of Russian spontaneous speech // *Proceedings of LREC*, 2016. Pp. 1949–1954.
 43. Kawahara T. et al. Synchrony in prosodic and linguistic features between backchannels and preceding utterances in attentive listening // *2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, IEEE, 2015. Pp. 392-395.
 44. Krivokapić J. Rhythm and convergence between speakers of American and Indian English // *Laboratory Phonology*, 2013. Vol. 4. N 1. Pp. 39-65.
 45. Levitan R., Hirschberg J. Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions // *Twelfth Annual Conference of the International Speech Communication Association*, 2011.
 46. Menshikova A., Kocharov D., Kachkovskaia T. Phonetic Entrainment in Cooperative Dialogues: A Case of Russian // *Proc. Interspeech 2020*, 2020. Pp. 4148-4152.
 47. Namy L. L., Nygaard L. C., Sauerteig D. Gender differences in vocal accommodation: The role of perception // *Journal of Language and Social Psychology*, 2002. Vol. 21. N 4. Pp. 422-432.
 48. Nenkova A., Gravano A., Hirschberg J. High frequency word entrainment in spoken dialogue // *Proceedings of ACL-08: HLT, Short Papers*, 2008. Pp. 169-172.

49. Pardo J. S. et al. Phonetic convergence in college roommates // *Journal of Phonetics*, 2012. Vol. 40. N 1. Pp. 190-197.
50. Pardo J. S. Expressing oneself in conversational interaction // *Expressing oneself/expressing one's self: Communication, cognition, language, and identity*, 2010. Pp. 183-196.
51. Pardo J. S. On phonetic convergence during conversational interaction // *The Journal of the Acoustical Society of America*, 2006. Vol. 119. N 4. Pp. 2382-2393.
52. Pickering M. J., Garrod S. Toward a mechanistic psychology of dialogue // *Behavioral and brain sciences*, 2004. Vol. 27. N 2. Pp. 169-190.
53. San Segundo E., Yang J. Formant dynamics of Spanish vocalic sequences in related speakers: A forensic-voice-comparison investigation // *Journal of Phonetics*, 2019. Vol. 75. Pp. 1-26.
54. Savino M. et al. Prosodic convergence in Italian game dialogues // *ExLing*, 2016. Pp. 151-154
55. Schweitzer A., Lewandowski N., Duran D. Social Attractiveness in Dialogs // *INTERSPEECH*, 2017. Pp. 2243-2247.
56. Soliz J., Giles H. Communication accommodation theory: A situated framework for relational, family, and intergroup dynamics // *Engaging theories in interpersonal communication: Multiple perspectives*, 2015. Pp. 157-170.
57. Street R. L., Giles H. Speech accommodation theory: A social cognitive approach to language and speech behavior // *Social cognition and communication*, 1982. Vol. 193226. Pp. 193-226.
58. Thakerar J. N., Giles H., Cheshire J. Psychological and linguistic parameters of speech accommodation theory // *Advances in the social psychology of language*, 1982. Vol. 205. Pp. 205-255.
59. Volskaya, N. and Kachkovskaia, T. Prosodic annotation in the new corpus of Russian spontaneous speech CoRuSS // *Proceedings of Speech Prosody*, 2016. Pp. 917-921.