

ИСПОЛЬЗОВАНИЕ ПРОГРАММНЫХ ПАКЕТОВ ОБРАБОТКИ ЗВУЧАЩЕЙ РЕЧИ В ИНСТРУМЕНТАЛЬНО- ФОНЕТИЧЕСКИХ ИССЛЕДОВАНИЯХ (методика анализа эксплозивной фазы немецких смычных взрывных согласных)

В области инструментальных исследований речи за последнее десятилетие фактически произошла революция, приведшая к полному вытеснению традиционных приборов, использовавшихся с начала 20 столетия при анализе речи. Эра закопченной бумажной ленты, мареевских барабанчиков, кимографов, осциллографов, спектрографов, положившая начало освоению "видимой речи" (*Visible Speech*) [17; 20], завершилась бурным прорывом в области компьютерных технологий, сделавших речь не только "видимой", но и слышимой в исполнении синтезированных голосов, приветствующих нас в метро, телефоне и даже в образе искусственно улыбающейся *Ananova*¹, рассказывающей в Интернете последние новости на английском, немецком и французском языках.

¹ Ананова – Первый в мире диктор-аватар (анимированный виртуальный персонаж с полочеловеческой индивидуальностью), "рожденный" в Соединенном Королевстве 17 апреля 2000 г. компанией *Lernout & Hauspie*, специализирующейся на речевых технологиях. Подробнее см. www.ananova.com.

Традиционно, инструментальное исследование звуковых единиц любого уровня в заданных (отобранных) речевых условиях, осуществляемое в рамках эксперимента или методом наблюдения, состоит из осциллографического, интонографического, спектрального видов анализа, что позволяет получить объективное представление об объекте исследования. Применение компьютерных речевых технологий делает возможным в настоящее время провести указанные виды инструментального анализа одновременно, делая тем самым процедуру обработки данных менее трудоемкой, более точной и наглядной.

Важным этапом в подготовке инструментального исследования звучащей речи является предварительная компьютерная обработка речевых образцов, состоящая из ряда последовательных операций: оцифровка аналогового сигнала, выделение необходимых отрезков на осциллограмме, их сохранение в формате *.wav, обработка в программном пакете.

Оцифровка аналоговых сигналов, записанных на магнитную пленку, может производиться при помощи программы редактирования звуковых файлов, например, *Goldwave* путем присоединения источника этих сигналов к звуковой карте персонального компьютера. Шестнадцатирядные звуковые файлы, полученные в результате оцифровки с частотой дискретизации 22 кГц, сохраняются на твердом диске в формате *.wav, поддерживаемом большинством программ анализа звучащей речи. Данная частота дискретизации наиболее подходит для исследований единиц сегментного уровня, в частности, согласных звуков, спектр которых, как правило, охватывает высокие частоты (8000-10000 Гц). При этом, частота дискретизации согласно теореме Котельникова должна превышать максимальную частоту звука хотя бы в два раза [1; 2].

Из полученных в результате оцифровки больших звуковых файлов размером более 2 Мбайт выделяются отрезки, содержащие исследуемые звуковые объекты. Эти отрезки сохраняются в виде отдельных файлов в соответствующих директориях.

Далее оцифрованный материал обрабатывается в одной из программ, например, PRAAT, WinCECIL, Spectrogram, Speech Analysis Tools, Sal

и другие. Наиболее функциональной и удобной в использовании представляется PRAAT 4.0, позволяющая, среди прочего, выводить на экран компьютера осциллограмму выделенного отрезка, спектрограмму шириной анализа 5 мс и частотой 10 кГц в окне Хамминг, контуры пределов изменения интенсивности в диапазоне от 40 до 100 дБ, цифровые значения энергетических характеристик сегментов с использованием начальных и конечных данных выделенного сегмента [14].

Особенностью используемого программного обеспечения является возможность получить не только визуальное изображение отсегментированного участка с различной степенью увеличения, но и прослушать его в разных режимах скорости, которые задаются пользователем. Искомый сегмент выделяется при помощи двух курсоров (основного и вспомогательного) и прослушивается в одном из выбранных режимов. Для точности на экран монитора выводится одновременно до трех окон с изображением одного и того же отсегментированного участка в виде осциллограмм, динамических спектрограмм, огибающих контуров интенсивности. Полученные результаты измерений представляются в виде электронных таблиц с целью дальнейшей обработки данных.

Важным моментом процесса сегментации является разграничение пловиза (или любого другого согласного) и предшествующего гласного, особенно, если идет обработка сегментов слитной речи. На осциллограмме за начало пловиза принимается точка пересечения нуля (zero-crossing point) после последнего глоттального импульса (glottal pulse) предшествующего гласного (рис.1) [16; 22]. На спектрограмме гласный идентифицируется за счет падения (повышения) уровня спектральной энергии в области 1000 Гц. Частотные и энергетические параметры рассчитываются в различных точках дискретизации выделенных сегментов.

Для выполнения спектрального анализа программа PRAAT 4.0 использует быстрые преобразования Фурье (БПФ). БПФ обычно определяются количеством входящих исходных точек, используемых в каждом расчете, которые всегда являются кратными двум (1024). Частот-

ное разрешение спектрограммы представляет собой цифровую частоту выборки, разделенную по частотам на количество исходных точек БПФ. Чем больше количество исходных точек БПФ, тем точнее частотное разрешение спектрограммы [11, 149-157].

Использование программного обеспечения позволяет, безусловно, получить высококачественное графическое изображение выделенного сегмента с обширной информацией о характере речевого сигнала. Важным, однако, является умение правильно расшифровать полученное изображение и извлечь из него достоверную информацию о проанализированном отрезке. Для этого существенным является понимание следующих моментов.

Во-первых, при проведении инструментального анализа, связанного с получением информации о единицах сегментного уровня, необходимо учитывать **систему взаимоотношений артикуляторных и акустических изменений**: качественные модификации речевого сигнала, описываемые в терминах спектрального анализа, являются вторичными по отношению к артикуляторным изменениям, носящим первичный характер [9, 65-66].

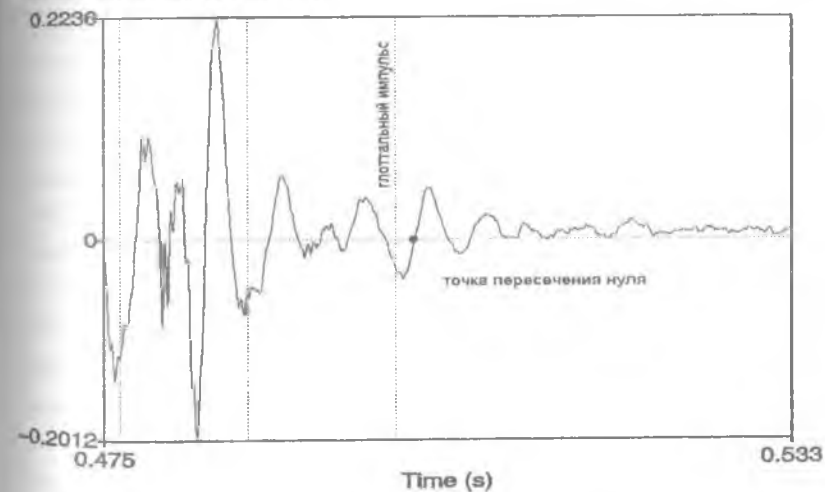


Рис.1. Точка начала сегментации предстыкового глухого пловиза [t] на осциллограмме

Однако, рассчитывать на прямое соответствие артикуляторных и акустических показателей в процессе речепроизводства возможно только при произнесении отдельно взятых звуков, что само по себе является искусственной ситуацией и возможно только на первичном этапе, когда исследователя интересует "чистый" вариант артикуляторно-акустической структуры звука, то есть его базовый вид. В реальных условиях производства слитной речи возникает явление артикуляторного полиморфизма, т.е. при различных положениях органов речи в пределах определенной зоны достигается схожий акустический эффект. Другими словами, в определенный момент говорения происходит соединение (наложение) нескольких артикуляторных жестов при единстве акустической структуры полученного сигнала [12, 235]. С другой стороны, на характеристику сигнала влияет позиция артикулируемого речевого сегмента в высказывании, то есть здесь проявляется зависимость его сугубо артикуляторно-акустических характеристик от просодической организации единиц более высокого порядка (предложение, текст) [8; 10]. Следовательно, перед сегментатором возникает сложная задача разграничения в одном сегменте артикуляторных, акустических и просодических характеристик и, соответственно, их корректного анализа.

Далее, обрабатывая информацию о единицах сегментного уровня, необходимо учитывать феномен сочетаемости *непрерывности и дискретности звуковых единиц* [3, 125-126]. Воспринимаемая дискретность слов, фраз в реальной естественной речи отсутствует. Она является лишь результатом предшествующего языкового опыта говорящего и реализуется через знание системы того или иного языка путем использования языкового контекста и ситуации. Следовательно, между звуками существуют только фонологические границы, а не фонетические (артикуляторные, акустические). Такой подход к пониманию процессов, лежащих в основе речеобразования, выводит на решение задач, связанных, во-первых, с правильным отбором материала исследования, его инструментальной обработкой, условиями записи, качеством пленки и т.д, то есть с тем, что может повлиять на физи-

ческие характеристики сигнала и исказить реальную лингвистическую картину (трактовку). Другой блок проблем основывается на разработке четких критериев сегментации речевого потока и многоуровневом подходе, т.е. на учете всех сегментных, супraseгментных и экстралингвистических факторов, влияющих на акустические характеристики звуков [5]. Учет вышеизложенных условий организации исследования позволит использовать информацию о полученных сегментах и их признаках для дальнейшего анализа естественной речи и синтеза речи, приближенной к натуральной.

Таким образом, приступая к изучению акустической структуры речевого сигнала, необходимо принимать во внимание ряд факторов: лингвистические, а именно, артикуляторный (коартикуляторный), акустический, просодический, а также экстралингвистические, среди которых чаще всего принято выделять гендерный, т.е. зависимость звукового образа от типа голоса (женский, мужской), эмоциональный (коннотативный), а также технический (технологический), связанный с условиями корректной обработки аналогового сигнала, его оцифровки, качества магнитной пленки и условий записи.

В настоящей статье рассматривается методика организации инструментального фонетического исследования при помощи программного пакета обработки звука Praat 4.0, разработанного в Амстердамском институте фонетических исследований и успешно апробированного в ряде работ, проводимых в Лаборатории экспериментальной фонетики Одесского национального университета им. И.И.Мечникова. В качестве объекта исследования выбрана взрывная фаза немецких смычных взрывных согласных p, t, k, находящихся на стыках слов и выделенных из слитной речи в виде звукосочетаний. Объем выборки составляет 1400 единиц. Подробнее о методике отбора материала и классификации стыковых позиций см. [7]. Исследование характеристик взрывной фазы проводилось с учетом следующих условий: типа стыка, т.е. гоморганности/гетероганности соединения, наличия паузы в звукосочетании, положения относительно ударного слога в фонетическом слове, а также типа голоса (женский/мужской).

Провести детальное исследование такого микросегмента как explosive phase, длительность которой в среднем не превышает 12 мс, стало возможным именно благодаря применению компьютерной обработки речевого сигнала, позволившей получить данные темпоральных, частотных и энергетических характеристик фазы взрыва. Расчет показателей производился исходя из значений длительности explosive phase (мс), частотного распределения энергии (Гц) и максимумов концентрации энергии взрыва в спектре (дБ).

Идентификация степени интенсивности взрыва на спектрограмме может осуществляться простым визуальным способом – на основе соотношения черно-серых цветовых характеристик сигнала [5, 26; 18, 24]. Параллельно с наблюдением цветовых характеристик взрыва в соответствующем рабочем окне программы ведется измерение интенсивности энергии взрыва в дБ. Так, максимальное количество энергии характеризуется уровнем интенсивности от 90 дБ и выше, что находит выражение в наличии областей черного цвета и легко идентифицируется визуально. Уровень интенсивности взрыва от 75 до 90 дБ может быть охарактеризован как средний. На спектрограмме такие области выражены темно-серым цветом. Падение интенсивности ниже 75 дБ характеризует взрыв как слабый, что будет выражаться в размытости его контуров или невозможности его сегментации в целом.

Замер спектральных пиков explosive phase производится ручным способом путем выведения рабочего курсора в точку максимального отклонения амплитудной кривой на выделенном сегменте (рис.2). В этой точке делается спектральный срез, содержащий необходимую информацию о распределении энергии на выделенном сегменте (рис.3). Точность измерений достигается за счет многократного увеличения (расширения) сегмента в рабочем окне. В таблицы вносятся начальные и конечные показатели длительности фазы взрыва, что позволяет при повторных замерах в случае необходимости произвести измерения с максимальной точностью.

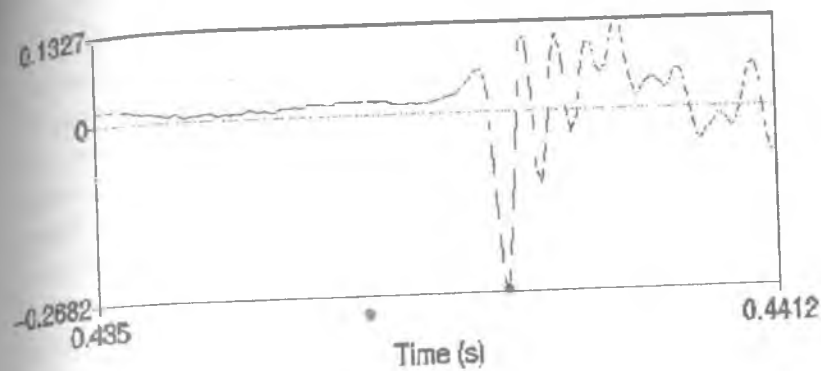


Рис. 2. Точка максимального отклонения амплитуды explosive phase [t]

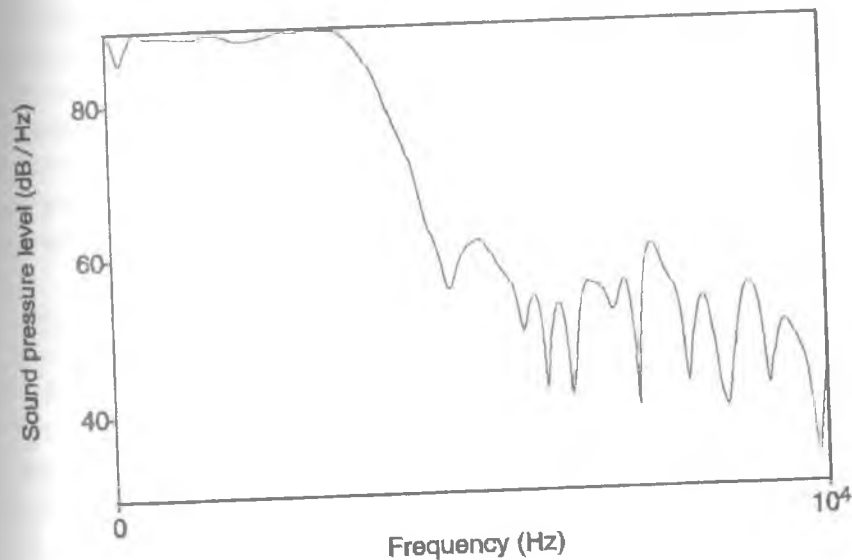


Рис.3. Спектральный срез в точке максимального отклонения амплитуды explosive phase [t]

При идентификации эксплозивной фазы на спектрограмме важными являются исходные сведения о характере и структуре взрыва, отражающие, прежде всего, *место образования* исследуемого пловива. Так, *билабиальный* пловив характеризуется самой слабой интенсивностью и наиболее короткой фазой разрешения смычки. Энергия взрыва концентрируется в области низких частот (500-1000 Гц). Взрыв *альвеолярного* пловива характеризуется прерывистой структурой и отсутствием многократного разрешения смычки. Максимальная концентрация энергии взрыва расположена в области высоких частот (свыше 4000 Гц). *Велярный* пловив характеризуется самым интенсивным взрывом с концентрацией энергии в области 1500-4000 Гц, а также наличием многократного разрешения фазы смычки, возникающего за счет большей площади контакта спинки языка с нёбом.

Следовательно, структура и уровень интенсивности взрыва могут служить надежным критерием сегментации и распознавания глухих пловивов в слитной речи, особенно в том случае, если аспирированный пловив находится в невокалическом окружении, т.е. там, где отсутствуют формантные переходы [19].

При изучении спектральной картины смычных взрывных согласных звуков принимался во внимание ряд *дифференциально* значимых акустических противопоставлений, необходимых для описания структуры акустического сигнала исследуемых звуков в консонантных звуко сочетаниях на стыке слов.

Соотношение между объемом резонирующей полости перед сужением (смыканием) и после него выражается в противопоставлении *компактный/диффузный*. Компактные звуки характеризуются преобладанием центрально локализованного формантного района. У диффузных звуков энергия резонансных частот рассеяна по большому отрезку шкалы. К группе компактных звуков относятся велярные [k]/[g], а к группе диффузных альвеолярные [t]/[d] и билабиальные [p]/[b].

Противопоставление *низкий/высокий* акустически выражается в уровне концентрации энергии в спектре. Если энергия сосредоточена в низких частотах спектра, звук характеризуется как низкий (гравис-

ый), если соответственно энергия концентрируется в высоких частотах – как высокий (акутный). В артикуляторном плане такое противопоставление связано с объемом и расчлененностью резонирующей полости. К низким относятся билабиальные и альвеолярные пловивы, к высоким – велярные. Следовательно, акустическое противопоставление *низкий/высокий* соответствует изменению ротового резонатора по горизонтальной оси. Дополнительно в момент образования низких звуков происходит сужение фаринкса, тогда как высокие звуки образуются при расширенной глотке [6, 17].

Вышеуказанные противопоставления являются определяющими при описании акустического параметра *эксплозивности* смычных взрывных согласных: смещение характерных областей взрыва в нехарактерные для описываемого звука будут свидетельствовать об определенной степени модификаций согласного по месту образования в зависимости от ряда комбинаторных и просодических факторов.

В результате исследования *эксплозивности* немецких смычных взрывных звуков было установлено, что длительность эксплозивной фазы предстыковых глухих пловивов составляет в среднем 10-12 мс и является средней величиной для всех звуков этой группы, независимо от их места образования ($p > 0.7$). Изменения эксплозивных характеристик определяется не только просодией речевого высказывания, но и взаимоотношениями между сегментами самого пловива.

Как известно, уровень интенсивности взрыва определяется *местом образования* пловива, что выражается в следующей зависимости: **чем больше объем резонатора, тем более выражена интенсивность взрыва**. Энергия взрыва билабиального согласного концентрируется в области низких частот (500-1000 Гц), альвеолярного пловива в области высоких частот (свыше 4000 Гц), велярного пловива характеризуется самым интенсивным взрывом с концентрацией энергии в области 1500-4000 Гц [19].

Исследование частотного распределения энергии эксплозивной фазы глухих пловивов показало, что в слитной речи происходит сдвиг характерных энергетических областей взрыва, определяемых *местом* образо-

вания пловива. Так, у билабиального [p] отмечается повышение верхней частотной границы до уровня 1860 Гц, при этом нижняя граница области взрыва остается стабильной и не превышает, как правило, 700 Гц. Незначительная рассеянность энергии по шкале частот (максимальное значение 630 Гц) говорит, скорее, о компактном характере пловива, чем о диффузном. Велярный предстыковый [k] сохраняет свою характеристику компактного звука. Ширина полосы взрыва также распространяется по незначительному отрезку шкалы и составляет максимально 1850 Гц. Частотная область концентрации энергии взрыва в основном касается средних частот, при этом верхняя граница взрыва не превышает 3500 Гц. Альвеолярный пловив [t] соответствует своей характеристике диффузного звука, поскольку его энергия рассеяна по значительному отрезку шкалы и достигает максимальной ширины 3470 Гц. Частотные области концентрации энергии взрыва альвеолярного глухого пловива характеризуются нестабильностью, что выражается в смещении нижней границы взрыва в область низких частот (500 Гц).

Приведенные ниже в табл. 1 данные частотного распределения энергии взрыва глухих предстыковых звуков позволяют охарактеризовать билабиальный смычный взрывной [p] как грависный (низкий) звук, а альвеолярный [t] и велярный [k] звуки как акутные (высокие).

Таблица 1

Частотное распределение энергии взрывной фазы глухих смычных взрывных согласных [p], [t], [k] (n = 51) (Гц)

[p]		[t]		[k]	
Нижняя граница	Верхняя граница	Нижняя граница	Верхняя граница	Нижняя граница	Верхняя граница
50	350	4700	5390	2380	3400
950	1860	4020	4760	800	2650
650	1100	3090	4680	400	2160
1600	2230	220	2060	1500	3230
20	660	300	1250	90	930
50	790	850	1300	80	800
400	1070	1000	1340	1740	2340
700	1000	1260	2700	1150	3000

[p]		[t]		[k]	
Нижняя граница	Верхняя граница	Нижняя граница	Верхняя граница	Нижняя граница	Верхняя граница
420	1020	3600	4680	1000	1660
510	900	2900	4300	710	1020
780	1210	1900	3150	490	1940
80	490	3650	4000	890	1090
80	910	1200	1900	2060	2500
600	1400	450	1500	470	2160
810	1200	600	1900	1540	2000
750	1000	2020	2240	1290	3100
60	590	4320	4860	2080	3700
500	1045	2122	3059	1098	2216

В отличие от аспиративной фазы длительность взрывной фазы не зависит от длительности интерсегментной паузы в звукосочетании ($p > 0.3$). Средние значения длительности взрывной фазы предстыковых глухих пловивов с паузой составляют 10,7 мс, без паузы – 10,1 мс.

Наличие/отсутствие паузы приводит к существенным различиям в энергетических характеристиках взрыва, а именно в показателях максимумов интенсивности всей взрывной фазы, а также в спектральных срезах, проводимых в точках максимального отклонения амплитуды в зонах концентрации энергии взрыва. Так, при наличии пауз среднеарифметические показатели пиков интенсивности составляют 83,2 дБ, при отсутствии паузы внутри звукосочетания – 74,3 дБ ($p < 0,001$).

Влияние интерсегментной паузы на реализацию энергетических характеристик фазы взрыва может быть рассмотрено на конкретных примерах.

Так, в сочетании слов Macht teilen (рис.4), где длительность интерсегментной паузы равна 50 мс, максимальная интенсивность взрыва в спектре предстыкового пловива [t] составляет 98,6 дБ. Энергия взрыва концентрируется в области средних частот (3090-4680 Гц). На рис. 5 представлен спектральный срез в точке максимального отклонения амплитуды в фазе взрыва предстыкового пловива [t].

В гетероорганичном сочетании пловизов [tb] на стыке слов *nacht Bodenfrost* пауза отсутствует (рис. 6). Вследствие этого максимальное значение интенсивности взрыва в спектре предстыкового пловиза [t] равно 64,2 дБ (рис. 7). Визуально это выражается отсутствием областей черного цвета в области взрыва.

Значительно отличаются также показатели уровня интенсивности всей эксплозивной фазы. В среднем с паузой они составляют 62,05 дБ, без паузы – 60,2 дБ ($p < 0,02$).

Положение пловиза в ударном слоге не является значимым для длительности и интенсивности эксплозивной фазы. Средние значения длительности эксплозивной фазы предстыкового пловиза в ударном слоге равны 11,7 мс, в безударном – 9,0 мс ($p > 0,3$). Средние значения интенсивности взрыва в спектре в ударном слоге составляют 84,3 дБ, в безударном – 82,1 дБ ($p > 0,2$).

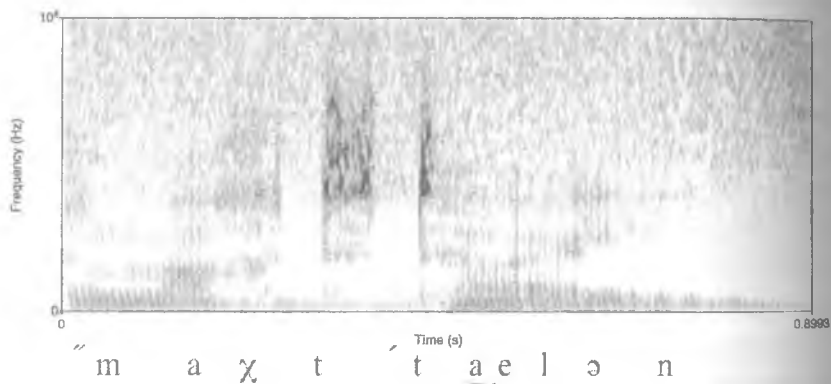


Рис. 4. Спектрограмма словосочетания *Macht teilen*

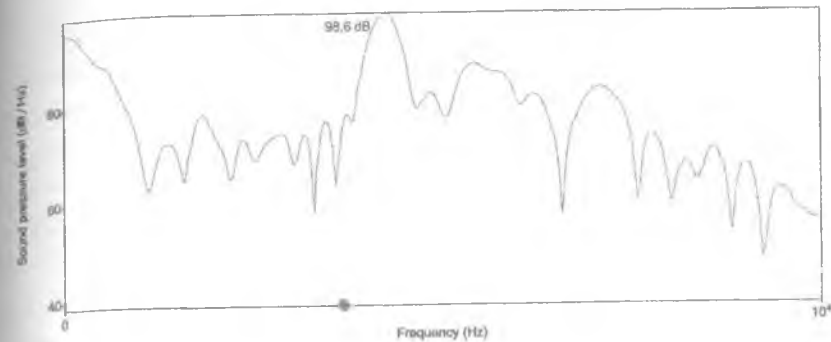


Рис. 5. Спектральный срез в области взрыва предстыкового пловиза [t] в сочетании *Macht teilen*

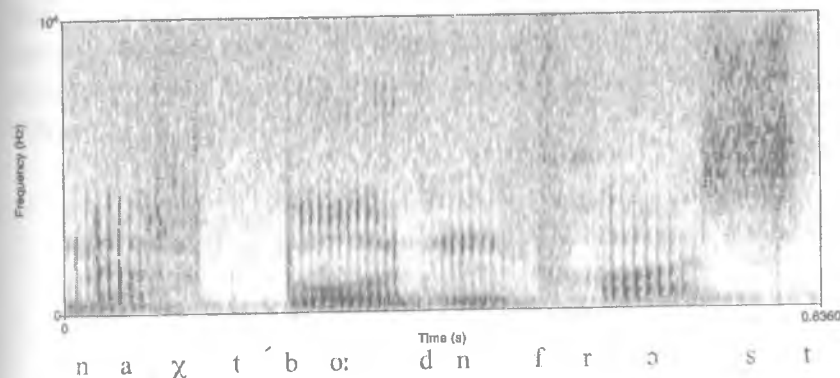


Рис. 6. Спектрограмма словосочетания *nacht Bodenfrost*

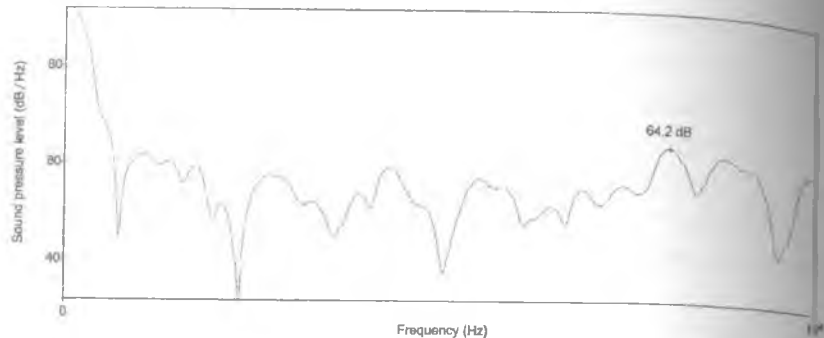


Рис.7. Спектральный срез в области взрыва предстыкового пловива [t] на стыке слов *nacht Bodenrost*

На сегментном уровне обнаружена взаимосвязь между интенсивностью фазы смычки и интенсивностью эксплозивной фазы ($r = 0,58$; $p < 0,001$), выражающаяся в следующей зависимости: чем больше интенсивность смычки, а значит меньше напряженность смыкательного движения, тем больше интенсивность взрыва (рис. 8). Данная корреляция может объяснить механизм ленизации фортисных пловивов.

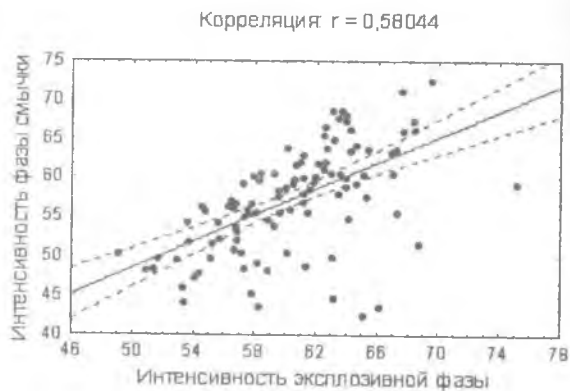


Рис. 8. График корреляции между интенсивностью фазы смычки и интенсивностью эксплозивной фазы

Отсутствие паузы в звукосочетании приводит не только к ослаблению энергетического уровня звуков, но и к различным видам звуковых модификаций, одним из которых является ассимиляция по способу образования. При определении основных причин звуковых изменений подобного рода нужно исходить, прежде всего, из гоморганности/гетероорганности сочетаний согласных на стыке слов, а также типа последующего застыкового согласного.

В гоморганных сочетаниях пловив + пловив [td], пловив + фрикатив [tz], пловив + сонорный [tn] в случае отсутствия интерсегментной паузы эксплозивная фаза первого звука может слиться с импловивной фазой второго звука.

В первом случае гоморганное звукосочетание трансформируется в один консонант с максимально длительной фазой смычки. На рис. 9 приведен образец спектрограммы сочетания слов *nicht teilnehmen*, где на стыке слов идентифицируется один глухой аспирированный пловив [t] с удлинненной фазой смычки (101 мс), низкая интенсивность которой (42,2 дБ) говорит о значительной напряженности смыкательного артикуляторного жеста. Это также отражается в показателях мощности звука, она равна 348,81 мкПа².

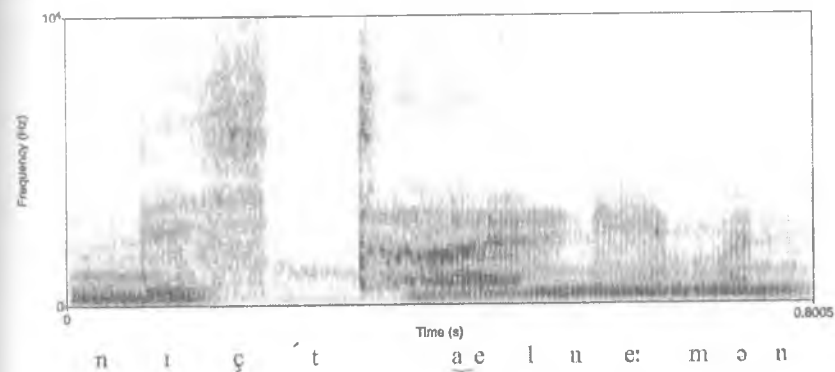


Рис. 9. Спектрограмма словосочетания *nicht teilnehmen*

Если в качестве застыкового звука выступает гоморганный звонкий шловив [d], то при отсутствии интерсегментной паузы также происходит реализация одного компонента стыка, застыкового полувзвонкого [d]. На рис.10 на стыке слов *Kandidat des* идентифицируется реализация полувзвонкого [d] с удлинненной фазой смычки (64 мс) при общей длительности звука 76 мс. Длительность эксплозивной фазы составляет 12 мс и характеризуется концентрацией энергии в области первой форманты. Максимальная интенсивность взрыва в спектре равна 83, 2 дБ.

В гоморганном сочетании шловив + фрикатив в шловиве образуется плавный переход к щели, что приводит к замене этого звукосочетания на аффрикату (рис.11). На спектрограмме стыка слов *hat seinen* идентифицируется слияние предстыкового альвеолярного шловива [t] с застыковым альвеолярным фрикативом [z], что приводит к образованию аффрикаты. Второй компонент стыка представляет собой оглушенную шелевую [s] с характерной областью высокочастотных шумов (3500-6400 Гц) и отсутствующей "голосовой линией" в области 500 Гц, типичным признаком звонких звуков. Компоненты стыкового звукосочетания реализуются со значительной артикуляторной напряженностью. Мощность звуков [t] и [s] равна, соответственно, 118,73 мкПа² и 90,11 мкПа².

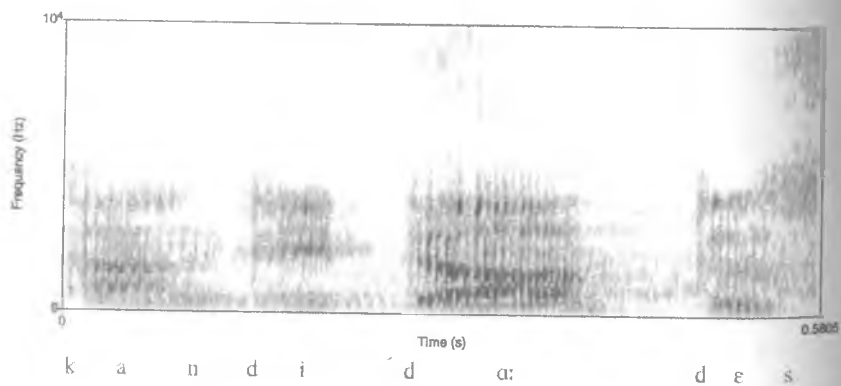


Рис.10. Спектрограмма словосочетания *Kandidat des*

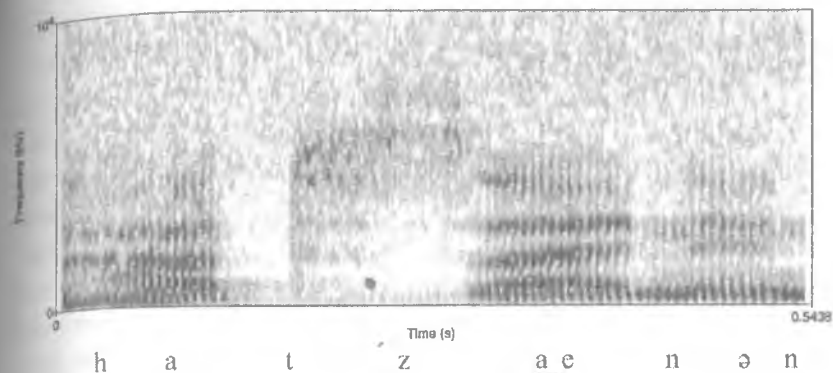


Рис. 11. Спектрограмма словосочетания *hat seinen*

В сочетаниях шловив + аффриката в случае отсутствия интерсегментной паузы происходит слияние шловива [t] с первым компонентом аффрикаты [ts], т.е. гоморганным звуком. На спектрограмме *Bereitschaft zur* в стыковом звукосочетании [tts] идентифицируются только удлинненная фаза смычки (55 мс) шловива [t] с последующей зоной высокочастотного шума, соответствующей сегменту глухого шелевого [s] (рис.12). Контурсы взрыва не определяют, следовательно, первым компонентом аффрикаты является дезэксплозивированный смычный звук. Подобная модификация является результатом спирализации, т.е. быстрого преобразования фазы смычки в щель, вызванной ослаблением смыкательного артикуляторного жеста (59 дБ).

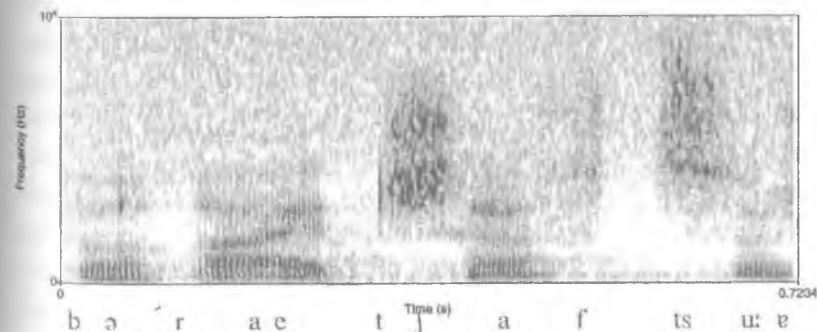


Рис.12. Спектрограмма словосочетания *Bereitschaft zur*

В гетероорганных сочетаниях пловив + постальвеолярный фрикатив значительное сокращение длительности эксплозивной фазы и ослабление интенсивности взрыва в данном виде стыка возникает в результате наложения турбулентного шума застыкового фрикатива с концентрацией энергии в области высоких частот. Здесь прослеживается прямая зависимость длительности аспирированной фазы и интенсивности взрыва от артикуляторной (акустической) глубины контактного щелевого: чем глубже по шкале частот расположена нижняя граница фрикативного шума, тем слабее аспирация контактного пловива [19]. В контактном положении предстыкового пловива с застыковым постальвеолярным фрикативом [ʃ], например, на стыке слов *mit Schnee*, обладающим максимальным объемом концентрации энергии в области 2500-7000 Гц, отсутствие аспирации фиксируется во всех без исключения случаях (рис. 13).

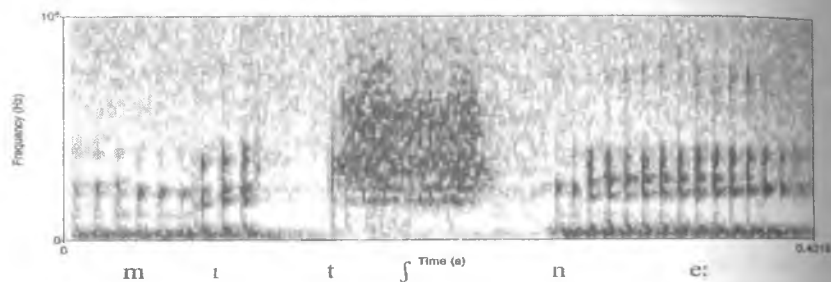


Рис.13. Спектрограмма словосочетания *mit Schnee*

Стыковые звуко сочетания пловив + глоттальный [h] представляют особую трудность при расшифровке спектрограмм, постольку здесь сложно провести сегментацию аспирированного пловива и глоттального фрикатива, особенно если в звуко сочетании отсутствует пауза. Как правило, в таких стыках происходит наложение областей глоттального [h] на фазу аспирации предстыкового пловива. В сегменте глоттального [h] на стыке слов *geführt hat* четко прослеживаются формантные переходы к последующему гласному (рис. 14), которые протекают параллельно. В результате сильного влияния последующего вокалического сегмента застыковый глоттальный фрикатив реализуется в звонкой форме.

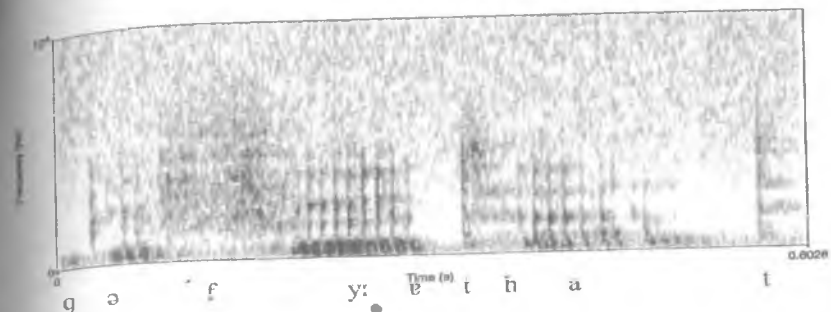


Рис. 14. Спектрограмма словосочетания *geführt hat*

Гоморганное сочетание пловива и сонорного характеризуется отсутствием аспирированной и эксплозивной фаз, что приводит только к реализации смычки и ее удлинению до 70-100 мс, например, на стыке слов *bestimmt nicht* (рис. 15). В этом случае предстыковый согласный становится импловзивным. Одновременно в данном виде стыка происходит замена орального взрыва фаукальным, образующимся при переходе к сонорному согласному вследствие опускания нёбной занавески [24, 179]. Смычка при переходе от [p] к [m], или от [t] к [n] не прерывается, начало артикуляции сонорного характеризуется лишь отрывом нёбной занавески от задней стенки носоглотки, что создает эффект фаукального (носового) взрыва [23, 26].

В сочетании гетероорганных пловивов при сохранении взрыва предстыкового согласного смычка застыкового пловива может осуществляться до взрыва предстыкового звука. В этом случае существенное значение имеет порядок следования согласных. Если в сочетании предшествует пловив более глубокого образования (например, велярный [k] перед альвеолярными [t/d] или билабиальными [p/b]), то его взрыв сводится только к отделению активного органа от пассивного, за которым не следует выхода воздуха. Такой взрыв практически не дает акустического эффекта и характеризуется минимальной длительностью (3-5 мс) и слабой интенсивностью (55-60 дБ). Уровень концентрации энергии взрыва перемещается в область низких частот (1000-1600Гц) и по уровню интенсивности не превышает 75-80 дБ, что

Сокращение указанной длительности паузы внутри звуко сочетания [tg] на стыке слов *Tod gerissen* приводит к реализации неаспирированного предстыкового пловиза [t] при сохранении взрывной фазы с уровнем интенсивности взрыва 85-90 дБ и концентрацией энергии в области высоких частот (3000-5000 Гц), что является характерным для альвеолярного пловиза (рис. 18). В целом взрыв будет ослабленным, так как количество выходящего при этом воздуха невелико, оно ограничено небольшим запасом, который содержится в пространстве между передней и задней смычкой [4, 229].

Сегментный анализ гетероорганых сочетаний пловизов на стыке слов показал, что реализация фаз эксплозии и аспирации предстыковых глухих пловизов в сочетании с последующим гетероорганым пловизом возможна только при наличии интерсегментной паузы длительностью более 80 мс. В случае сокращения длительности интерсегментной паузы до 40-65 мс возможна реализация неаспирированного пловиза с достаточно структурированным взрывом, при этом уровень концентрации энергии взрывной фазы может смещаться в нетипичные для пловиза частотные уровни.

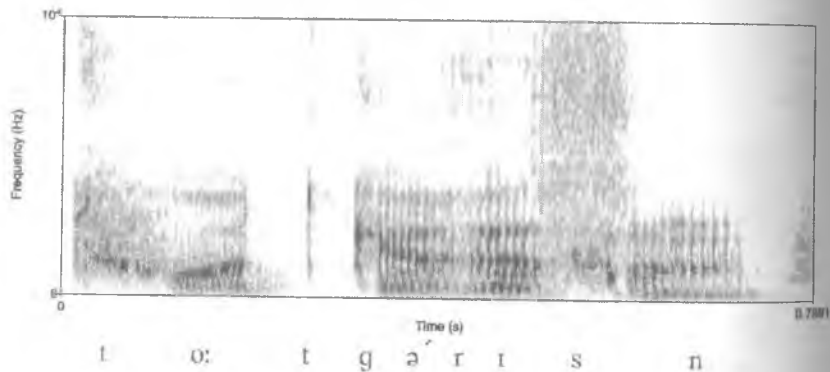


Рис. 18. Спектрограмма словосочетания *Tod gerissen*

Последовательное сравнение средних значений длительности и интенсивности взрывной фазы предстыковых смычных звуков показало отсутствие существенных различий в зависимости от их положения

в ударном слоге. Длительность взрывной фазы в акцентной позиции составляет 11,7 мс, в безударном слоге – 9,0 мс ($p > 0,3$). Средние показатели интенсивности взрыва в спектре в ударном слоге составляют 84,3 дБ, в безударной позиции – 82,1 дБ.

Также положение звуко сочетаний на стыках слов внутри и на границах ФС является не существенным при определении их различий. Длительность взрывной фазы смычных предстыковых звуков на границах слов в рамках ФС составляет 11,2 мс, на стыках слов, образующих границы фонетических слов – 10,3 мс ($p > 0,2$). Средние показатели интенсивности взрывной фазы в спектре составляют 84,6 дБ, если звуко сочетание находится на стыке слов *внутри* ФС, 81,2 дБ для пловизов, являющихся пограничными *на стыке* ФС ($p > 0,2$).

Несущественными являются различия по показателям средних величин длительности интерсегментных пауз и взрывной фазы глухих предстыковых пловизов в зависимости от гендерного фактора. В целом, средние показатели длительности пауз являются различными и составляют у женщин 60,2 мс и у мужчин 54,2 мс, но эти различия не являются статистически значимыми ($p > 0,1$). Средние значения длительности взрывной фазы в реализации дикторов-женщин составляют 10,35 мс, дикторов-мужчин – 10,15 мс ($p > 0,7$).

Сравнение средних показателей интенсивности взрывной фазы у женских и мужских голосов так же показало несущественность полученных различий ($p > 0,2$). Средние значения интенсивности взрывной фазы для женских голосов составляют 59,2 дБ, для мужских – 61,3 дБ.

В результате исследования взрывной фазы предстыковых глухих смычных звуков были определены причины, вызывающие модификации данного сегмента пловизов. Прежде всего, это высокий уровень антиципирующей коартикуляции, возникающий только при определенной последовательности звуков на границах слов, где артикуляторные жесты последующего звука доминируют над артикуляторными жестами предыдущего. Эти изменения обнаружены как в сочетаниях звуков, сходных по способу образования, но различающихся

ся по месту артикуляции, так и в сочетаниях звуков, различных по способу образования. В гетероорганных сочетаниях пловизов будет доминировать билабиальная смычка над велярной и альвеолярной, если билабиальный звук будет находиться в застыковой позиции. В сочетаниях с фрикативными звуками, в той же последовательности будет обладать влияние застыкового щелевого звука, поскольку он обладает большими энергетическими характеристиками.

Нереализация эксплозивной фазы связана также с ослаблением напряженности смыкательного жеста, о чем свидетельствуют результаты полученной корреляции между значениями интенсивности фазы смычки и взрыва. Результатом ослабления смыкательного жеста является спирализация смычных взрывных пловизов. Смещение характерных для каждого из видов пловизов зон концентрации энергии взрыва также является следствием сильных коартикуляторных процессов, возникающих в слитной речи при связывании слов в единое высказывание. Сохранение структуры отдельных сегментов звуков, в частности предстыковых глухих пловизов возможно только при наличии длительных интерсегментных пауз, хотя и это условие не может быть однозначным, поскольку взаимовлияние звуков в слитном произнесении не прекращается в полной мере даже при этом условии.

Столь точное исследование такого микросегмента как эксплозивная фаза немецких смычных взрывных согласных стало возможным благодаря применению современных программных пакетов обработки речи, которые, несомненно, будут использоваться и в дальнейших инструментальных исследованиях речи, связанных, прежде всего, с просодическим анализом более крупных единиц, таких как сверхфразовое единство, текст.

Литература

1. Анерт В., Райхардт В. Основы техники звукоусиления. – М.: Радио и связь, 1984. – 320 с.
2. Гоноровский И.С. Радио-технические цепи и сигналы. – М.: Советское радио, 1967. – 439 с.

3. Дукельский Н.И. Принципы сегментации речевого потока. – М.: Л.: Изд-во АН СССР (Ленинградское отделение), 1962. – 138 с.
4. Зиндер Л.Р. Общая фонетика. – М.: Высшая школа, 1979. – 312 с.
5. Зиндер Л.Р. Фонема и восприятие // Proc. of the 6th Int. Congr. Phon Sci. (Prague, 7-13 Sept., 1967) – Prague, 1970. – P. 1070-1073.
6. Златоустова Л.В., Потапова Р.К., Трунин-Донской В.Н. Общая и прикладная фонетика. М.: Изд-во МГУ, 1986. – 304 с.
7. Петлюченко Н.В. Реализация консонантных сочетаний на стыке лексических единиц (инструментально-фонетическое исследование на материале речи дикторов радио и телевидения Германии): Дис. ...канд. филол. наук: 10.02.04. – Одесса, 1999. – 230 с.
8. Потапова Р.К. Просодические характеристики макросегментации слитной речи // Экспериментальная фонетика. – М.: Изд-во МГУ, 1989. – С. 125-146.
9. Потапова Р.К. Слоговая фонетика германских языков. – М.: Высшая школа, 1986. – 144 с.
10. Потапова Р.К., Гордеева Т.А. К вопросу о пограничных сигналах в современном немецком языке // Вопросы языкознания. – М., 1998. – № 2. – С. 118-128.
11. Применение цифровой обработки сигналов. / Под. ред. Э.Оппенгейма. – М.: Мир, 1980. – 552 с.
12. Реформатский А.А. Введение в языковедение. / Под ред. В.А. Виноградова. – М., 1998. – 536 с.
13. Щерба Л.В. Языковая система и речевая деятельность. – Л.: Наука, 1974. – 428 с.
14. Boersma P. Praat, a system for doing phonetics by computer. – <http://fonsg3.let.uva.nl/>.
15. Fowler C., Saltzman E. Coordination and coarticulation in speech production // Language and Speech 36, 1993. – P. 171-195.
16. Giovanardi M., Di Benedetto M. G. Acoustic analysis of singleton and geminate fricatives in Italian // The European Student Journal of

17. Koenig, W., Dunn, H.K., Lacy, L.Y. The sound spectrograph // *JASA* – 1946. – Vol. 18. – P.19-49.
18. Kohler K. Einführung in die Phonetik des Deutschen. – Berlin: Erich Schmidt, 1977. – 251 S.
19. Machelett K., Tillmann H.G. Das Lesen von Sonogrammen. – München. 1996. <http://www.phonetik.uni-muenchen.de/SGL>.
20. Potter, R.K., Kopp, G.A., Green, H. Visible Speech. – New York, 1947.
21. Saltzman E., Munhall K. A dynamical approach to gestural patterning in speech production // *Ecological Psychology* 1, 1989. – P. 333-382.
22. Son – van R.J.J.H., Pols L.C.W. A comparison between the acoustics of vowel and consonant reduction // *Proceedings* Vol. 20. – Amsterdam: Institute for Phonetic Sciences, University of Amsterdam, 1996. – P.13-25.
23. Tillmann H.G., Schiel F. Akustische Phonetik. – München, 1998. – 233 S.
24. Wängler H. H. Grundriss einer Phonetik des Deutschen. – Marburg: N.G.Elwent, 1983. – 256 S.