ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ
1971г. вып.44

УДК 62-5:621.391.

СИНТЕЗ РУССКОЙ РЕЧИ ПО ПРАВИЛАМ

М.Деркач, Н.Загоруйко, Й.Лилиенкранц,
С.Паули, Г.Фант

1. Общие принципы синтеза речевых образцов

В настоящей работе изложены методика и результаты формантного синтеза русских слогов, слов и некоторых фраз по правилам, когда сообщение задается последовательностью фонем и его синтез контролируется электронно-вычислительной машиной. Работа была выполнена на синтезаторе *OVE*-Ш в лаборатории преобразования речи в Королевском технологическом институте в Стокгольме */ STL /*.

Были синтезированы следующие фонемы русского языка:
гласные -[у, о, а, э, ы, и],
глухие щелевые согласные -[с, ш, ф, х] и их звонкие пары - [з, ж, в],
аффикаты -[ц, ч],
глухие взрывные согласные -[п, т, к]и их звонкие пары -[б,д,г].

Синтезатор *OVE*-Ш, разработанный в лаборатории преобразования речи, отличается от его предшественника *OVE*-П [1] тем, что он управляется электронно-вычислительной машиной в цифровой форме [2, 3, 4] .

На рис. 1 приведена общая схема аппаратуры, используемой для синтеза речи. Центром управления синтезом является компютер *CDC*-1700 с оперативной памятью 8 тысяч 16-битовых слов и временем обращения к памяти - 1,1 мксек. К нему прилагается внешняя память на дисках и общий преобразователь, который обеспечивает связь между машиной и аналоговым лабораторным оборудованием, а также обеспечивает аналоговый выход на контрольный осциллоскоп. Пульт управления снабжен печатающей машинкой, световым пером и соответствующими регуляторами с цифровыми дешифраторами. Естественный и синтетический речевые образцы могут быть непосредственно сравнены на экране контрольного осциллоскопа с использованием 51-канального спектрального анализатора, имеющего цифровой выход, с получением, если нужно, функции рассогласования.

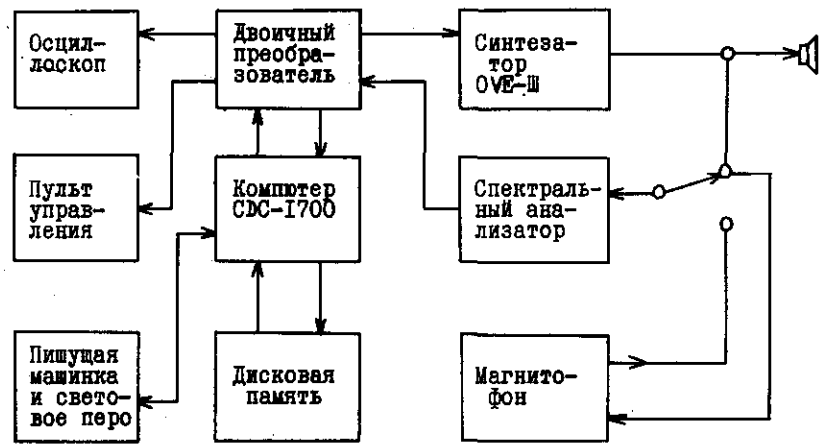


Рис.1. Общая схема аппаратуры для синтеза речи.

На рис.2 приведена блок-схема самого синтезатора. Система *OVE*-Ш составлена соответственно из 3-х отдельных последовательных ветвей, соединенных общим выводом. Первая из них - ветвь *F* - служит для задания конфигурации полости рта и определяется формантными частотами *F1*, *F2*, *F3* и *F4* с соответствующими ширинами полос. Ширина полосы, а также соответствующие постоянные времени сглаживающих фильтров, не контролируются в ходе непосредственного синтеза, а задаются предварительно. Формантная ветвь *F* может быть связана с источником голоса ИГ или шума ИШ или с тем и другим одновременно. В первом случае уровень звука управляется блоком управления амплитудой источника голоса АО. Это имеет место, например, при синтезе гласных. В том случае, когда ветвь *F* соединяется с источником шума, его уровень регулируется блоком управления амплитудой аспиративного шума АН в режиме образования придыхательных звуков (места управления по-

казаны широкими стрелками).

Вторая, фрикативная, ветвь, охватывающая два резонанса К1 и К2 и один антирезонанс К0, предназначена для формирования спектра сегментов фрикативного шума. Его интенсивность регулируется блоком управления уровня фрикативного шума АС.

Третья, назальная, ветвь составлена из назального резонанса FN и предназначена главным образом для воспроизведения назальных согласных или назализованных гласных, а блок AN предназначен для управления уровнем назальной системы.

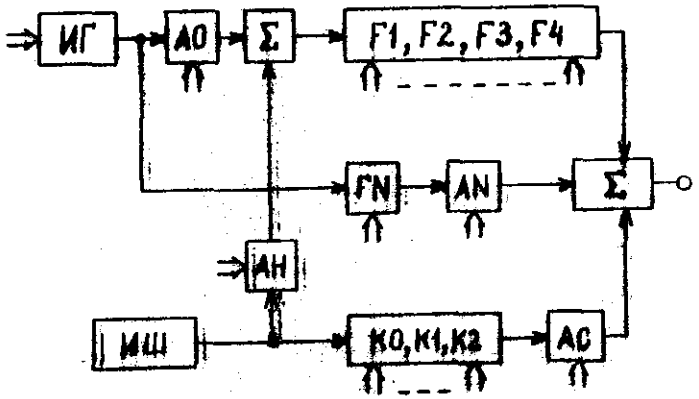


Рис. 2. Блок-схема конечного аналогового устройства синтезатора OVE - III.

Указанная организация управляемых параметров хорошо согласуется с развиваемой Г.Фантом акустической теорией специфика - ции сегментов согласно их фонетической природе, отражающей их образование [5, 6] . Следует подчеркнуть, что здесь относительные уровни формант не контролируются независимо, а функционально вытекают из частотного расположения задаваемых формант.

Идея фонемного синтеза состоит в том, чтобы обеспечить слитную разборчивую речь, задаваемую пофонемно печатанием последовательности буквенных символов с пульта управления, независимо от взаимосочетаний фонем в различных текстах. Поскольку синтезируемое сообщение задается как последовательность фонем, ис - ходная и принципиально важная часть работ по синтезу заключается в формировании в памяти машины библиотеки фонем. Каждый фонемный образец в библиотеке представляет собой набор управляе-

мых параметров, которые задаются ступенчато с соответствующими временными индексами и указанием абсолютной длительности для каждого из них. При составлении сочетаний фонем эти параметры стыкуются, сохраняя последовательность их поступления во вре - мени, а соответствующие фильтры сглаживают задаваемые ступен - чатые изменения, обеспечивая плавность акустического выхода и сохраняя специфику переходных процессов. Процедура ступенчатого управления синтезом речи исходит из концепции инерционности речедвигательного аппарата, управляемого дискретными импульсными командами нервной системы.

Исследовательская работа при фонемном синтезе слитной речи заключается прежде всего в проверке значимости соответствующих параметров и их спектрально-временных изменений и в отборе тех из них, которые оказываются наиболее существенными для достижения этой цели. Методика синтеза под визуальным контролем на экране осциллоскопа позволяет многократно изменять отдельные параметры в фонемных образцах, непосредственно прослушивать их звучание в желаемых звукосочетаниях и производить их разносторонний акустический анализ.

2. Синтез гласных фонем

Формантные частоты гласных были измерены на спектрографическом материале русских ГСГ*) - слогов в произнесении одного диктора. Они приведены в таблице I. Уровень интенсивности А0 для гласных был установлен 24 дб, а длительность (DR) -160 мсек (8 20-ти миллисекундных единиц длительности). Указанная дли - тельность сохранялась и для согласных фонем.

Таблица I

Формантные частоты синтезируемых гласных (в гц)

	$F1$	$F2$	$F3$
У	380	750	2250
О	450	850	2310
А	900	1450	2400
Э	550	2000	2450
Ы	400	2000	2500
И	350	2310	2550

*) гласный - согласный - гласный

В таблице 2 приведены параметры гласных фонем так, как они сохранялись в фонемной библиотеке машины.

Таблица 2

Параметры гласных в библиотеке фонем

[у]:	DR	8		[э]:	DR	8	
	AO	0	24		AO	0	24
	AC	0	0		AC	0	0
	AH	0	0		AH	0	0
	AN	0	0		AN	0	0
	F1	-3	380		F1	-3	549
	F2	-3	749		F2	-3	2001
	F3	-3	2245		F3	-3	2448
[о]:	DR	8		[ы]:	DR	8	
	AO	0	24		AO	0	24
	AC	0	0		AC	0	0
	AH	0	0		AH	0	0
	AN	0	0		AN	0	0
	F1	-3	452		F1	-3	400
	F2	-3	847		F2	-3	2001
	F3	-3	2310		F3	-3	2502
[а]:	DR	8		[и]:	DR	8	
	AO	0	24		AO	0	24
	AC	0	0		AC	0	0
	AH	0	0		AH	0	0
	AN	0	0		AN	0	0
	F1	-3	904		F1	-3	351
	F2	-3	1456		F2	-3	2310
	F3	-3	2396		F3	-3	2558

В приведенной таблице первый столбец цифр определяет время ступенчатого задания параметров в фонемном образце, а сами их значения указаны во втором столбце (для параметров уровня - в дб, для частотных параметров - в гц; временные координаты даны в единицах времени, каждая из которых равна 20 мсек). Следует обратить внимание на то, что частотные параметры задаются с опережением во времени так, чтобы они достигли номинальных значе-

ний к моменту включения источника голоса. Синтезированные гласные при прослушивании были вполне разборчивыми, но специальным артикуляционным испытаниям не подвергались.

3. Система согласных

Синтезируемые согласные были расклассифицированы соответственно основным фонетическим принципам их образования, как это показано в таблице 3. В этой классификации допущен ряд упрощений. Например, [ф] и [в] с точки зрения их образования являются губно-зубными (лабиодентальными), тогда как они помещены в группу лабиальных. Аналогично [т] и [д] являются альвео-дентальными, а в классификационной таблице они стоят в группе дентальных. Учитывая то обстоятельство, что такого рода некорректность описания для русского языка является только аллофонической, но не фонемной, описанная выше классификация согласных по месту их образования на лабиальные, дентальные, альвеолярные и велярные была сочтена приемлемой для синтеза русских сообщений.

Таблица 3

Фонетическая классификация, принятая при синтезе согласных

Место образования	Способ образования: щелевые		взрывные	
	Тип источника: глухие	звонкие	глухие	звонкие
губные	ф	в	п	б
зубные	с	з	т	д
альвеол.	ш	ж	-	-
велярные	х	-	к	г

В пределах каждой группы по месту артикуляции согласные отличались по способу их образования(сюда включается тип источников, а также временные характеристики их включения и выключения в синтезируемом звуке).

В таблице 4 приведены частотные характеристики синтезируемых групп согласных, полученные на спектрографическом материале одного диктора. В ней следует обратить внимание на то,что глу -

хие согласные звуки синтезировались не только с использованием фрикативного звена К, но также и ветви голосовых формант F в аспиративном режиме. Это было сделано невзирая на то, что нижние формантные области у ряда согласных или сильно подавлены по интенсивности, или же отсутствуют полностью, а поэтому не берутся в расчет во многих работах по синтезу. Оказалось, однако, что управление как формантным, так и фрикативным звеньями при синтезе согласных позволяет более надежно отобразить картину формантных переходов, а это имеет значение для обеспечения разборчивости. Более низкое значение $F1$ у группы взрывных согласных отражает полную смычку ротовой полости по сравнению с группой щелевых согласных. Следует подчеркнуть также особенности велярной группы, где числовые значения $F2$ не задавались самостоятельно, а заимствовались из окружающих гласных, отражая специфику весьма тесной корреляции этой группы с окружающими её фонемами.

Таблица 4

Частотные характеристики фонетических групп синтезируемых согласных (в гц)

Фонетическая группа	$F1$	$F2$	$F3$	К0	К1	К2
Губные						
щел. ф, в	250	800	2000	1700	1700	4000
взр. п, б	200					
Зубные						
щел. с, з	250	1800	2500	2500	5000	7700
взр. т, д	200					
Альвеолярные						
щел. ш, ж	250	1600	2500	1000	2000	3000
взр. -						
Велярные						
щел. х	500	определяется окружающими фонемами	1800	2500	1500	2500
взр. к, г	200					

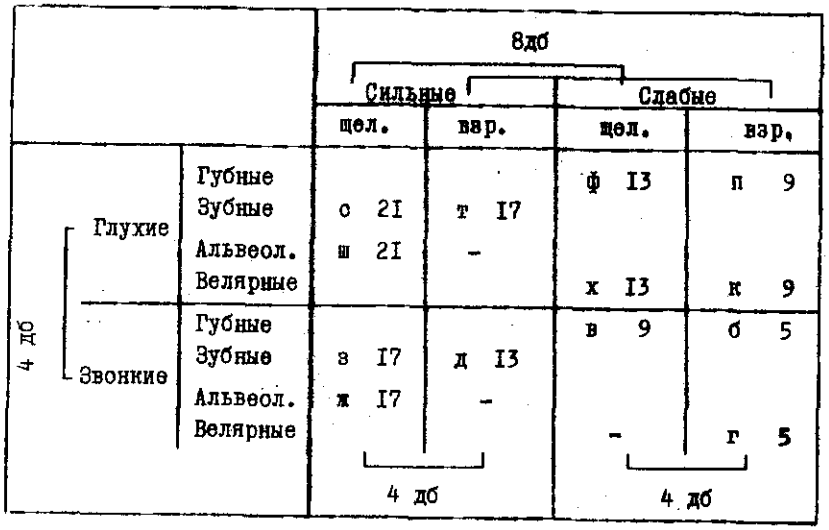
С учетом известных фактов о соотношении между уровнями фрикативного и аспиративного шумов АС и АН в естественной речи оно было жестко установлено как

АС : АН = 18 дб.

Что же касается более тонких различий по уровню между синтезируемыми согласными, то была предпринята попытка отразить различия типа напряженный-ненапряженный (*tense - lax*) и сильный-слабый (*strong - weak*) в терминологии дифференциальных признаков Р.Якобсона, Г.Фанта и М.Халле [7]. Эти различия были отнесены только к уровню фрикативного шума АС, как это показано в таблице 5.

Таблица 5

Дифференциальные различия в уровне фрикативного шума АС для групп синтезируемых согласных (в дб)



Наконец, признаки, относящиеся к отличиям по способу образования синтезированных согласных. Группу взрывных согласных отличало от группы щелевых введение акустической паузы в начале взрывного звука, которая у звонких согласных заменялась низкочастотной фонацией, задаваемой соответствующими параметрами назальной ветви AN и FN. Сам взрыв занимал конечные 40 мсек (2 единицы времени), из них параметры фрикативного шума, несущие информацию о спектре самого взрыва, длились первые 20 мсек, тогда как последние 20 мсек были заняты только аспиративным шумом, отражающим конфигурацию голосового тракта в формантном описании.

4. Синтез целевых согласных

Синтез целевых согласных был основан на управлении частотными и амплитудными параметрами фрикативного звена (К1, К2, К0, АС) и аспиративного звена (*F1*, *F2*, *F3*, АН) синтезатора, а в группе звонких согласных – также и голосовым источником (АО) и полосой низкочастотной фонации (*AN*, *FN*).

У группы глухих целевых согласных были выключены все источники гармонических колебаний (АО = 0[*]; А = 0) и включены все источники шума (АС ≠ 0; АН ≠ 0). Включение и выключение источников осуществлялось плавным нарастанием их интенсивности. Уровень фрикативного шума у [с] , [ш] устанавливался на 8 дб выше, чем у [ф],[х] . Этот признак, как показали мингографические исследования, статистически достоверен.

Были введены также отличия в уровень аспиративного шума АН для разных целевых согласных. Наиболее низкий уровень АН (13дб) был установлен для дентальной и лабиальной групп ([ф], [с]), для которых цель расположена на дистантном конце голосовой трубы. По мере приближения места положения цели к месту главного резонанса голосового тракта уровень аспиративного шума возрастал и для альвеолярного [ш] он составлял 21 дб, а для велярного [х] – 29 дб. Как следствие этого, форманты голосового тракта были усилены и отчетливо видны у велярной группы и ослаблены в случае дентальной и лабиальной групп.

Звонкие целевые согласные [в] , [з], [ж] отличались от глухих целевых двумя свойствами. Первое из них состояло в включении полосы низкочастотной фонации с частотой *FN* = 250 гц. Второе отличие относилось к уровню фрикативного шума АС :

АС (глух.цел.) : АС (зв. цел.) = 4 дб.

В таблицах 6 и 7 приведены соответствующие параметры синтезируемых глухих и звонких целевых согласных в машинной библиотеке фонем.

В каждой диаде чисел, разделенных большим интервалом, первое число указывает момент времени, а второе – значение параметра в этот момент времени в соответствующих единицах – дб – для параметров уровня и гц–для частотных параметров. Данные, приведенные в таблицах 6 и 7, отражают общие принципы синтеза глухих и звонких целевых согласных, описанные выше. Дополнительно

[*] Фактически АО = 1 дб, так как при АО = 0 конструкция *OVE*-II не позволяет включить источник аспиративного шума АН .

к ним следует обратить внимание на то, что глухие целевые имеют включенный голосовой источник АО в начальные 20 мсек их звучания, что противоречит принципу их образования. Это было допущено с целью продлить время перехода формант от гласного к согласному в слогах типа ГСГ, однако для общего случая описания целевых согласных этот прием следует признать неудовлетворительным.

На рис. 3 приведены результаты артикуляционных испытаний восприятия синтезированных глухих и звонких целевых согласных в слогах типа [а – согл – а] и [а – согл – у] , проведенные на 12 нетренированных аудиторах, обеспечивших 100 ответов для каждой фонемы.

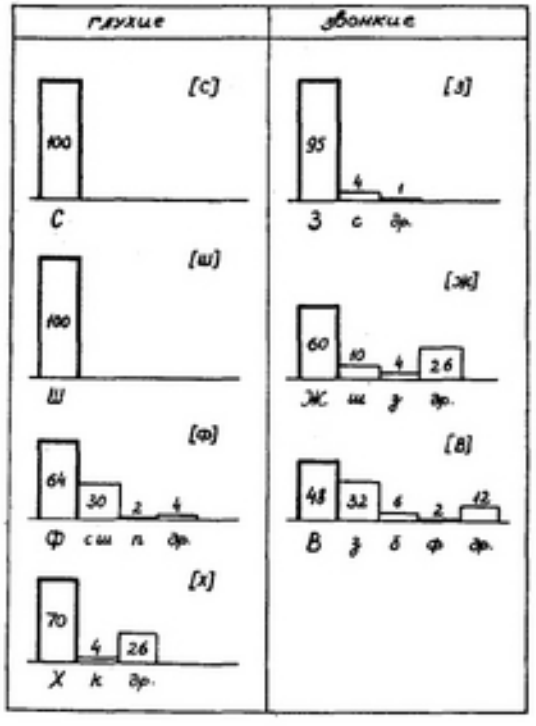


Рис. 3. Опознаваемость глухих и звонких целевых согласных в синтетических слогах типа Г С Г (в процентах).

Таблица 6

Параметры глухих щелевых согласных в библиотеке фонем

[С] :	DR	8					
	AO	0	I3	I	I		
	AC	0	I3	I	2I	7	I3
	AH	0	I3				
	AN	0	0				
	FI	0	250				
	F2	-I	I808				
	F3	-I	2502				
	KO	-5	2500				
	KI	-5	5004				
	K2	-5	7720				
[Ш] :	DR	8					
	AO	0	I3	I	I		
	AC	0	I3	I	2I	7	I3
	AH	0	2I				
	AN	0	0				
	FI	0	250				
	F2	-I	I600				
	F3	-I	2502				
	KO	-5	I000				
	KI	-5	200I				
	K2	-5	2997				
[Ф] :	DR	8					
	AO	0	I3	I	I		
	AC	0	5	I	I3	7	5
	AH	0	I3				
	AN	0	0				
	FI	0	250				
	F2	-I	800				
	F3	-I	200I				
	KO	-5	I706				
	KI	-5	I706				
	K2	-5	4002				
[Х] :	DR	8					
	AO	0	I3	I	I		
	AC	0	5	I	I3	7	5
	AH	0	29				
	AN	0	0				
	FI	0	500				
	F3	-I	I808				
	KO	-5	2539				
	KI	-5	I498				
	K2	-5	2539				

Таблица 7

Параметры звонких щелевых согласных в библиотеке фонем

[З] :	DR	8					
	AO	0	I7	3	0	6	I7
	AC	0	9	I	I7	7	9
	AH	0	I3				
	AN	0	I3				
	FI	0	250				
	F2	-I	I808				
	F3	-I	2502				
	FN	-5	250				
	KO	-5	2500				
	KI	-5	5004				
	K2	-5	7720				
[Ж] :	DR	8					
	AO	0	I7	3	5	6	I7
	AC	0	9	I	I7	7	9
	AH	0	2I				
	AN	0	I3				
	FI	0	250				
	F2	-I	I600				
	F3	-I	2502				
	FN	-5	250				
	KO	-5	I000				
	KI	-5	200I				
	K2	-5	2997				
[В] :	DR	8					
	AO	0	I7	3	0	6	I7
	AC	0	I	I	9	7	I
	AH	0	I3				
	AN	0	I3				
	FI	0	250				
	F2	-I	800				
	F3	-I	200I				
	FN	-5	250				
	KO	-5	I706				
	KI	-5	I706				
	K2	-5	4002				

На рис. 4 в качестве примера приведены динамические спектрограммы синтетического (слева) и естественного (справа) слога [аса], а на рис. 5 – аналогичные спектрограммы для слога [аза].

Рис. 4. Спектрограммы синтетического и естественного слога [аса].

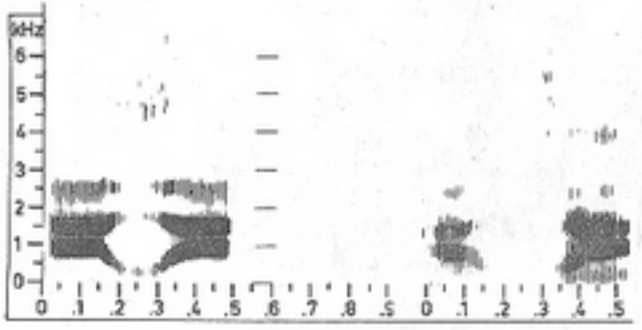


Рис. 5. Спектрограммы синтетического и естественного слога [аза].

5. Синтез взрывных согласных

В таблице 8 приведены параметры глухих взрывных согласных [п], [т], [к], а в таблице 9 – звонких взрывных [б], [д], [г] так, как они были представлены в фонемной библиотеке. Общие

Таблица 8
Параметры глухих взрывных согласных в библиотеке фонем

[П]:

DR	8					
A0	0	I3	I	0	6	I
AC	0	0	6	9	7	0
AH	0	0	6	I3		
AN	0	0				
F1	0	200				
F2	-I	800				
F3	-I	2001				
K0	-5	I706				
K1	-5	I706				
K2	-5	4002				

[Т]:

DR	8					
A0	0	I3	I	0	6	I
AC	0	0	6	I?	7	0
AH	0	0	6	I3		
AN	0	0				
F1	0	200				
F2	-I	I808				
F3	-I	2502				
K0	-5	2500				
K1	-5	5004				
K2	-5	7720				

[К]:

DR	8					
A0	0	I3	I	0	5	I
AC	0	0	5	9	7	0
AH	0	0	5	29		
AN	0	0				
F1	0	500				
F3	-I	I808				
K0	-5	2539				
K1	-5	I498				
K2	-5	2539				

Таблица 9
Параметры звонких взрывных согласных в библиотеке фонем

[Б]:

DR	8					
A0	0	I3	3	I	?	I3
AC	0	0	6	5	?	0
AH	0	0	6	I3		
AN	0	I3				
F1	0	200				
F2	-I	800				
F3	-I	2001				
FN	-5	200				
K0	-5	I706				
K1	-5	I706				
K2	-5	4002				

[Д]:

DR	8					
A0	0	I3	3	I	?	I3
AC	0	0	6	I3	?	0
AH	0	0	6	I3		
AN	0	I3				
F1	0	200				
F2	-I	I808				
F3	-I	2502				
FN	-5	200				
K0	-5	2500				
K1	-5	5004				
K2	-5	7720				

[Г]:

DR	8					
A0	0	I3	3	I	?	I3
AC	0	0	5	5	?	0
AH	0	0	5	29		
AN	0	I3				
F1	0	500				
F3	-I	I808				
FN	-5	200				
K0	-5	2539				
K1	-5	I498				
K2	-5	2539				

принципах их синтеза были описаны в параграфе 2. Так же, как и в случае щелевых, озвучивание глухих взрывных в первые 20 мсек их образования, направленное на подчеркивание характера формантных переходов в начале смычки, следует считать неудачным. Необходимо обратить внимание на то, что у звонких взрывных согласных положение частоты первой форманты $F1$, совпадающее с частотой фонации FN, занимает более низкое значение по сравне-

ния со щелевыми, что отражает полное смыкание рта в первом случае и образование щели во втором. В отличие от других взрывных фонема [к] имела большую длительность взрыва, приближающую её к аффрикатам. Это свойство было сохранено и у звонкого [г]. Интенсивность фрикативного шума АС у группы взрывных была ниже, чем у щелевых. Что же касается соотношения интенсивностей между глухими и звонкими взрывными, то оно было сохранено таким же, как и для группы щелевых согласных (см. таблицу 5).

Результаты артикуляционных испытаний синтезированных взрывных согласных в тех же условиях, которые были описаны в предыдущем параграфе, приведены на рисунке 6.

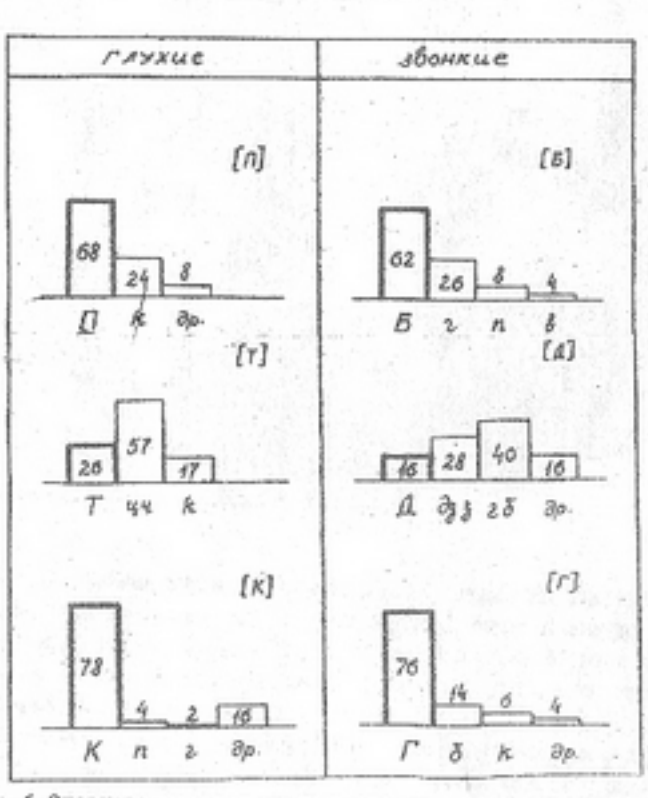


Рис.6.Опознаваемость глухих и звонких взрывных согласных в синтетических слогах типа Г С Г (в процентах)

В качестве примеров на рисунке 7 показаны спектрограммы синтетического и естественного слога [ата], а на рисунке 8 – такие же спектрограммы для слога [ада]. Обращает на себя внимание несоответствие спектральных характеристик взрыва у синтетических [т] и [д] естественному оригиналу, являющееся следствием того, что указанные фонемы синтезировались как зубные, а не как альвеоло – зубные.

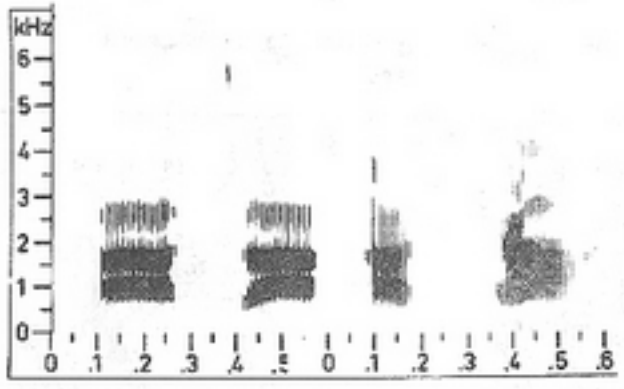


Рис. 7. Спектрограммы синтетического и естественного слога [ата]

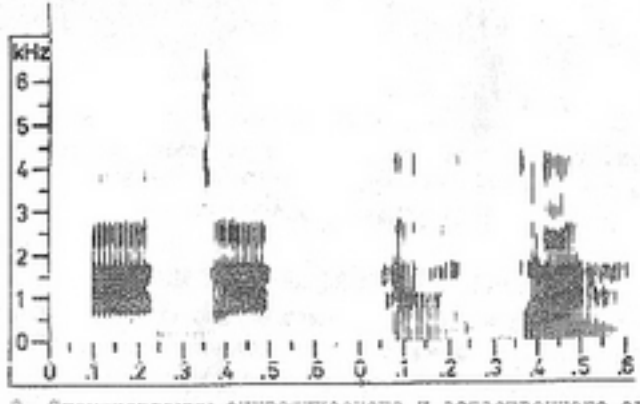


Рис. 8. Спектрограммы синтетического и естественного слога [ада]

6. Синтез аффрикат и фонемы [р]

Синтез аффрикат был осуществлен на базе соответствующих глухих щелевых согласных : [ц] - как укороченное [с] с паузой, занимающей первую половину его длительности, и [ч] - как укороченное таким же способом [ш] . Дрожащее [р] имело частотные параметры нейтрального гласного с быстрой сменой включения и выключения источника голоса АО. Библиотечные параметры этих фонем приведены в таблице IO .

Таблица IO

Параметры аффрикат и дрожащего [р] в библиотеке фонем

[Ц]:	DR	8						[Р]:	DR	8						
	AO	0	I3	I	0	4	I		AO	0	24	I	0	2	24	3 0
	AC	0	0	4	2I				AC	0	0					
	AH	0	0	4	I3				AH	0	0					
	AN	0	0						AN	0	0					
	F1	0	200						F1	0	500					
	F2	-I	I706						F2	-I	I498					
	F3	-I	2502						F3	-I	23I0					
	K0	-5	2500													
	KI	-5	5004													
	K2	-5	7720													
[Ч]:	DR	8														
	AO	0	I3	I	0	4	I									
	AC	0	0	4	2I											
	AH	0	0	4	I3											
	AN	0	0													
	F1	0	200													
	F2	-I	I600													
	F3	-I	2502													
	K0	-5	I000													
	KI	-5	200I													
	K2	-5	2997													

В артикуляционных испытаниях синтезированные аффрикаты [ц] и [ч] были правильно опознаны соответственно в 94% и 77% случаев. Фонема [р] на восприятие не испытывалась.

7. Синтез признака мягкости согласных

В настоящей работе особое внимание было уделено отображению правил коартикуляции фонем в синтезированном отрезке. Это означает, что частотные параметры формант последующих фонем номинально задавались ещё до окончания предыдущих. Вследствие инерционности исполнительных органов, моделируемой соответственными постоянными времени управляемых параметров, это приводило к появлению переходов, указывающих на направление сдвига формант. Эта информация оказывается важной для разборчивого синтеза почти всех фонемных групп. Она, в частности, необходима для надежного отображения мягкости согласных. В отдельном исследовании, специально посвященном этому вопросу, было показано,что для уверенного восприятия мягкости согласного в слоге типа ГСГ принципиально важно, чтобы μ - образная конфигурация голосового тракта, установившаяся в согласном, сохранялась до начала последующего гласного, придавая ему дифтонгоидный характер [8] . Для отображения этого феномена была введена команда мягкости (палатализации) [*] , которая предшествовала смягчаемому согласному :

[*]:	DR	0		
	F1	-I	290	9 290
	F2	-2	23I0	9 23I0
	F3	-2	2558	9 2558

Как видно, команда мягкости имеет нулевую длительность и поэтому сводится исключительно к управлению формантными частотами во времени так, что μ-образное расположение первых трех формант, даже если они исчезают на отрезке согласного в сонограмме, сохраняется до самого начала последующего гласного.Это обеспечивает появление полной картины формантных переходов в последующем гласном, напоминающей картину дифтонга, как это видно на рисунке 9, где приведены спектрограммы синтетического и естественного слога [ася].

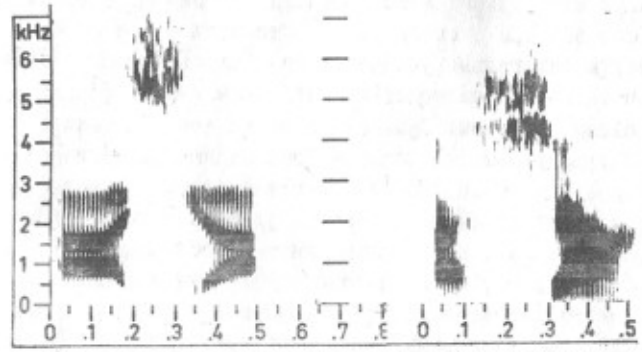


Рис.9. Спектрограммы синтетического и естественного слога [ася].

В артикуляционных испытаниях мягкость согласных воспринималась во всех без исключения синтезированных таким образом слогах. В пределах фонетических групп ошибок было в среднем больше, чем в группе твердых согласных. Лучше всего опознавалась группа мягких зубных согласных ([с'] - 100%, [з'] - 85%, [ц']- 88%, [т']- 84%, [д']- 88%), затем - группа мягких альвеолярных согласных ([н'] - 60%, [л'] - 45%, [ч'] - 80%); следующее место занимала группа мягких велярных согласных ([х'] - 44%,[к']- 28%, [г']- 40%) и самые худшие результаты были у группы мягких губных согласных ([ф'] - 28%, [в']- 0%, [п']- 9%, [б']- 0%). Следует при этом иметь в виду, что не все они реально существуют в естественной русской речи.

Средняя пофонемная разборчивость для всего материала синтезированных слогов составляла 60%. Это соответствует разборчивости на уровне фраз порядка 90% [9] .

8. Синтез фраз

Синтез фраз был осуществлен, в основном, с демонстрационной целью. С пульта управления печатались последовательности фонем с использованием интонационного рисунка, принцип которого состоял в следующем. Символы повышающейся интонации :

[↑]: D R 0
F 0 0 60

и понижающейся интонации :

[↓]: D R 0
F 0 0 160

управляли изменением частоты основного тона F 0 источника голоса, обладая при этом большой постоянной времени (400 мсек) , но которую желательно увеличить ещё больше, чтобы избежать монотонности голоса на участках между командами, управляющими интонационным контуром. Сущность замысла при введении интонационного контура фразы сводилась к "оживлению" синтетического голоса поочередным повышением и понижением основного тона в начале слов , а также в начале и в конце ударных гласных. Ударные гласные при этом имели полуторную длительность (символ [:] после гласного), а в случае логического ударения в слове - двойную длительность (двойная печать гласного). Ниже приводится образец записи фразы "Девушка, как тебя зовут ?" так, как она печаталась в процессе синтеза:

↑...↓ДЭ:↓ВУШКА↑ ...↓КА:К. ↑ КТЭБЭ↓А:↑. ↓ЗАВ↑УУ↓Т.

На рисунке 10 приведена спектрограмма этой синтетической фразы, а на рисунке 11 - она же в естественном произношении.

Речевой выход из синтезатора получался немедленно после окончания печати фразы.

Оценка качества синтезируемых фраз, по предварительным данным, дала удовлетворительные результаты.

9. Заключение

Как метод изучения акустических параметров речи, отражающих способ их образования и важных для восприятия, синтез является принципиально важным методом для исследования информационной природы речи. Принципы синтеза, изложенные в настоящей работе, достаточно упрощены и поэтому имеют ценность только в плане развития метода в последующем. В частности, результаты проведенной работы показали места введения необходимых коррекций в методику синтеза. Помимо целого ряда частных вопросов о самих параметрах, характеризующих отдельные фонемы, хочется отметить важность более корректного управления интенсивностью звуков в синтезируемом сообщении, а также текущего управления постоянными времени изменения используемых параметров в процессе синтеза в будущем. Представляется целесообразным снабдить синтезирующее устройство логической программой, позволяющей учитывать аллофонические тонкости и просодические факторы в синтезируемой речи. В более далекой перспективе синтез речи должен быть дополнен его текущим автоматическим распознаванием, производящим оценку корректности синтеза и вносящим необходимые исправления в цепи обратных связей. Разработка автоматов такого рода может оказаться полезной для исследования речевого процесса и для ряда прикладных задач.

Результаты проведенной работы вселяют оптимистическую уверенность в том, что речь может образовываться на пофонемной основе, а, следовательно, может и распознаваться с фонемных позиций. В этом плане синтез речи в управляемых автоматах является мощным экспериментальным методом непосредственной проверки значимости как отдельных акустико - фонетических параметров в речевом потоке для слухового восприятия речевой информации, так и для изучения принципов организации речевого процесса в целом. Всё это связывает проблемы автоматического распознавания, ана-

Рис. I0. Спектрограмма синтетической фразы "Девушка,как тебя зовут?"

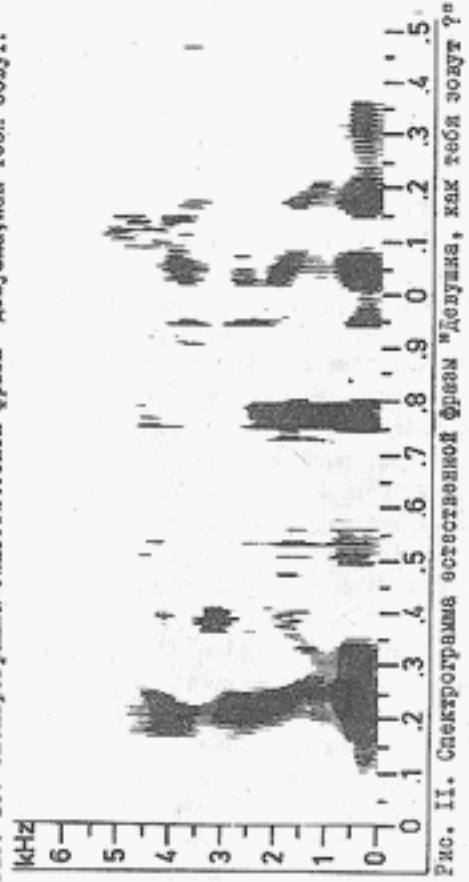


Рис. II. Спектрограмма естественной фразы "Девушка, как тебя зовут ?"

лиза и синтеза речи в единый комплекс проблем речевой коммуникации, приобретающий в настоящее время всё большее научное и практическое значение.

Литература

1. G.Fant a. J.Mártony. "Instrumentation for Parametric Synthesis (OVE- II)". Speech Transmiss. Lab., Stockholm,Quart. Progr.a.Status Rept., 2/1962, pp. 18-24.
2. J.Liljencrants. "The OVE-III Speech Synthesizer". Speech Transmiss Lab., Stockholm, Quart. Progr. a. Status Rept., 2-3/1967, pp. 76-81.
3. J.Liljencrants. "The OVE-III Speech Synthesizer". IEEE Transactions on Audio a. Electroacoustics, vol.AU-16,№1, March 1968, pp. 137-140.
4. J.Liljencrants. "Speech Synthesizer Control by Smoothed Step Functions." Speech Transmiss Lab., Stockholm, Quart.Progr. a. Status Rept., 4/1969, pp. 43-50.
5. Г.Фант. "Акустическая теория речеобразования". Перевод с англ. Изд-во "Наука", Москва, 1964.
6. Г.Фант. "Анализ и синтез речи". Перевод с англ. Изд-во "Наука", Новосибирск, 1970.
7. Р.Якобсон, Г.Фант и М.Халле. "Введение в анализ речи". Перевод с англ. В сб. "Новое в лингвистике", том.2,Москва, 1962.
8. M.Derksch, G.Fant. a. A. de Serpa-Leitão. "Phoneme Coarticulation in Russian Hard and Soft VCV-Utterances with Voiceless Fricatives". Speech Transmiss.Lab., Stockholm, Quart. Progr. a. Status Rept., 2-3/1970, pp. 1-7.
9. Н.Б.Покровский. "Расчет и измерение разборчивости речи". Связьиздат, 1962.

Поступила в редакцию
3.3.1971 г.