
**КОНФЕРЕНЦИИ
И СОВЕЩАНИЯ**

УДК 534.781

ДИАГНОСТИЧЕСКИЕ АРТИКУЛЯЦИОННЫЕ ТАБЛИЦЫ

© 2002 г. В. Г. Михайлов

Московский государственный университет
119899 Москва, Воробьевы горы
1ГК, тел. 939 32 56
E-mail: rfc-lsve@mtu-net.ru

В последние годы значительные успехи достигнуты в развитии инструментальных методов обобщенной оценки качества звучания и разборчивости на основе процедур моделирования слухового восприятия речи и измерений отношения сигнал/шум. Несмотря на ряд достоинств (быстрота измерений при низкой трудоемкости), эти методы не универсальны и по существу вторичны, так как для градуировки используют результаты субъективно-статистических измерений. Вместе с тем, некоторые специфические вопросы оценки качества, такие как диагностика причин отклонения качества звучания от нормативного, например акцентных признаков речи диктора или индивидуальных искажений голоса, находят свое решение с помощью психолингвистических методов. В статье рассматриваются диагностические артикуляционные таблицы в виде минимальных пар односложных слов DRT, соответствующие модели дифференциальных признаков Р. Якобсона, таблицы из многосложных четверок русских слов (метод выбора), неполных односложных слоговых таблиц _ГС/СГ_ (метод дописывания). Даются сравнительные оценки таблиц и рекомендации по их применению.

Качество звучащей речи определяется совокупностью признаков, характеризующих естественность, узнаваемость, разборчивость (понятность), громкость, наличие посторонних призывков и помех. Цель тестирования заключается в установлении степени соответствия качества звучания некоторым нормам, выявлении источников искажения и их устранения. Тестирование звучащей речи осуществляется в интересах телефонии (оценка качества передачи для тракта связи), педагогики (обучение произношению на иностранном языке), медицины (выявление дефектов речи и слуха), криминалистики (идентификация личности по голосу) и др. Методы тестирования делятся на субъективно-статистические (психолингвистические), выполняемые с участием дикторов и слушателей, и инструментальные, осуществляемые путем расчетов и автоматизированных измерений. Субъективно-статистические измерения прошли длительный путь применения и совершенствования. В развитие методов абонентской оценки качества передачи значительный вклад внесли Л.Р. Зиндер (1951), М.А. Сапожников (1959), Н.Б. Покровский (1962), Ю.С. Быков (1959), а также Н. Eletcher (1929), G. Fairbanks (1958), К. Kryter (1965). В последние годы опубликованы результаты ряда новых исследований проблемы тестирования [Р.К. Потапова (1986), Л.В. Златоустова (1986), Г.В. Вемян (1985), D. Pisoni (1989), W. Voiers (1983)].

Методы измерений регламентированы в России государственными стандартами. Стандартом

ГОСТ 50840-95, введенным в действие в 1997 году, предусмотрено измерение качества передачи, разборчивости и узнаваемости диктора по голосу в трактах речевой коммуникации [6]. Для тестирования применяются как методы измерения с использованием артикуляционных таблиц (слов, фраз), так и методы вербального описания качества звучания – методы мнений, парных сравнений и др. Словесные и слоговые артикуляционные измерения проводятся с использованием таблиц слов и слогов. Основное назначение таблиц слов – измерение разборчивости в трактах связи с предельно допустимым качеством в условиях действия сильных акустических шумов. Из-за ограниченного числа таблиц (всего 50) и высокой запоминаемости частое употребление этих таблиц недопустимо. Возможны также измерения с помощью таблиц фраз и двузначных цифр. Первый тип таблиц практически не применяется из-за присущей им большой избыточности и вследствие этого низкой чувствительности к искажениям в тракте связи. Так, разборчивость фраз при удовлетворительном качестве тракта равна 95–97%, хорошему – 97–99.5% и отличному – 99.5–100%. Цифровые таблицы используются в специфических условиях чрезвычайно низкой разборчивости (менее 25%), недопустимой для телефонной связи.

Между разными видами разборчивости установлена определенная зависимость [1]. Таблицы слогов обладают рядом недостатков: не отражены просодические параметры речи (контур ос-

новного тона, ритмическая структура и темп речи); спектральные характеристики звуков в составе слогов имеют утрированную форму; наконец, для получения необходимой величины коэффициента экспериментальной практики артикуляционной бригады ($z = 0.96-0.98$) требуется длительная тренировка, особенно при восприятии синтезированной речи [2].

В этом отношении рядом достоинств обладают словесные таблицы. Эти таблицы содержат слова с ритмическими структурами разного типа по числу слогов и месту ударения; обеспечивается естественность произнесения; требуемое время тренировки артикуляционной бригады слушателей (аудиторов) мало. Однако таблицы слов обладают значительной избыточностью, т.е. низкой чувствительностью к искажениям речевого сигнала. Результаты измерений разборчивости слов и слогов представляются в виде статистически усредненных оценок качества передачи без

описания лингвистических факторов, вызвавших искажения единиц речи. Этот недостаток призваны устранить диагностические артикуляционные таблицы.

Диагностические таблицы DRT. Известны диагностические артикуляционные таблицы минимальных пар односложных слов DRT – Diagnostic Rhyme Test, разработанные на основе модели дифференциальных признаков Р. Якобсона [4], таблицы из многосложных четверок русских слов (метод выбора) [5] и таблицы неполных слогов _ГС/СГ_ без начального/конечного согласного (метод дописывания) [6].

Диагностические испытательные таблицы DRT были разработаны в Кембриджской лаборатории ВВС США для оценки влияния искажений сигнала в тракте связи на различимость дифференциальных признаков фонем. Таблицы состоят из 112 пар односложных слов английского языка, например:

beam-peen	need-deed	feet-peat	seem-theme	peach-teach	she-see	*teal-keel
gin-chin	mitt-bit	fin-pin	chink-kink	bid-did	shift-sift	tilt-kilt
dent-tent	mend-bend	fend-pend	jest-guest	pest-test	yen-wen	pen-ken
vast-fast	nab-dab	fan-pan	champ-camp	bad-dad	cast-past	ram-yam

* Седьмая пара строки по выбору экспериментатора.

Пары слов различаются в порядке следования по одному из дифференциальных признаков английских согласных (Т.Н. минимальные пары): звонкий, носовой, непрерывный, напряженный, бемольный, компактный. Седьмая пара слов группы имеет признак по выбору экспериментатора. Всего в таблице 16 групп по 7 пар слов в каждой. Слова содержат гласные переднего ряда /i, e/ и гласные заднего ряда /u, o/_ – высокого и низкого подъема. Составители таблицы стремились выдержать принцип: пара слов – признак, хотя в некоторых случаях есть отклонения, например, согласные /k – p, g – b/ различаются по признаку “компактный”, а также по признаку “простой”.

При измерениях таблица (по одному слову из пары) повторяется 4 раза, что занимает 10 мин (по 1.3 с на слово). Признак пары слов для чтения, например, звонкий/глухой, выбирается в случайном порядке, повторяясь за испытание дважды. Слушатели отмечают переданное слово по полной отпечатанной таблице. В испытании участвует бригада из одного или двух дикторов с хорошо поставленным голосом и восьми слушателей с нормальным слухом. Различимость согласных, подсчитанная по формуле $d = 100(R - r)/T$, где R – число правильных отметок, r – число ошибочных отметок и $T = 448$ – число переданных слов, ока-

залась близка к разборчивости 256 фонетически балансированных односложных слов.

Таблицы DRT широко используются за рубежом для измерения разборчивости синтезированной речи. Заметность искажения согласных по различительным признакам включается в показатель абонентской приемлемости связи DAM (Diagnostic Acceptability Measure) [4]. Измерения разборчивости при воздействии помех и частотных искажений в тракте связи с полосой пропускания 200–4000 Гц показали, что шум по-разному влияет на различимость дифференциальных признаков фонем. Наиболее чувствительными к шуму оказались признаки прерванный, простой, компактный и наименее – звонкий, носовой. Отмечается также неодинаковое воздействие шума на противопоставляемые признаки, например, признак дизный оказался более чувствительным к шуму, чем признак простой. Как и при воздействии шума, сильное ограничение верхних частот сильно влияет на признаки прерванный, простой, компактный и менее заметно – на признаки звонкий, носовой. Вместе с тем, различимость согласных при искажении остается достаточно высокой, что обусловлено достаточно устойчивым слуховым восприятием лексических значений слов в минимальных парах. По данным многих измерений, величина различимости близка к 70–95%.

Таблица 1

Разборчивость (%) для испытательных таблиц			
Диктор	Слоговые	Диагностические (словесные)	Диагностические (парные звуко сочетания)
Мужчина	85.6	99.2	94.6
Женщина	83.4	98.8	98.0
Среднее	84.1	98.8	97.5
	84.4	98.9	96.7

Диагностические таблицы выбора. Попытки создания диагностических таблиц на материале русского языка сталкиваются с рядом трудностей. Система дифференциальных признаков, используемая для противопоставления согласных английского языка, не может быть без изменений перенесена на русский, не только из-за отличия своих признаков. В русском языке число односложных лексических единиц ограничено. Помимо гласных букв а, ы, э, у, о существуют йотированные варианты я, е, ю, ё, использование некоторых из представленных гласных (например, ы, э после ряда согласных может образовывать сочетания, не свойственные русскому языку (например, кы, гю, хе и др.). В ряде случаев невозможно подобрать минимальные пары из односложных слов с различием согласных по одному признаку. Поэтому был предложен так называемый метод выбора [5]. Метод состоит в передаче по испытуемому тракту трех таблиц по 27 слов в каждой. На приеме операторы (не менее трех) должны отметить переданное слово по таблице из 27 групп фонетических сходных (близких по звучанию) слов по четыре слова в каждой группе, например, слово бант из группы бланк, план, бант, банк или слово греть из группы угореть, гореть, болеть, балет и т.д. Мерой качества (разборчивости) служит среднее число ошибок на одну таблицу: две ошибки – высший класс, 5 ошибок – I класс, 8 ошибок – II класс и, наконец, 11 ошибок – III класс. Всего было составлено 10 таблиц выбора.

Число вариантов чтения одной таблицы может быть достаточно большим. При чтении со скоростью одно слово за три секунды время передачи одной таблицы равно 1.5 мин, одного сеанса из 30 таблиц – около 60 мин (с интервалом между

каждыми таблицами около – 6–8 мин). Специальных требований к тренировке бригады не предъявляется. Перед испытанием операторы прослушивают несколько таблиц для тренировки. Если среднее число ошибок от таблицы к таблице не уменьшается, то тренировка считается законченной.

В лаборатории прикладной лингвистики МГУ были разработаны диагностические таблицы для русского языка в виде таблиц парных звуко сочетаний СГС, в основе которых лежит противопоставление начальных или конечных согласных по одному из различительных признаков русских согласных (шумность/сонорность, место образования, способ образования, звонкость/глухость, твердость/мягкость) [7]. При испытании слушатель отмечает по таблице минимальных пар воспринятое звуко сочетание.

Оценка таблиц DRT и таблиц выбора. С целью сравнения чувствительности описанных диагностических таблиц к искажениям речевого сигнала были проведены артикуляционные измерения разборчивости синтезированной речи на модели вокодера LPC – 10 со скоростью передачи – 2.4 кбит/с. В измерениях участвовала бригада из трех дикторов и четырех слушателей. Полученные результаты приведены в табл. 1.

Видно, что по сравнению со слоговыми таблицами по ГОСТ Р 50840-95 таблицы выбора имеют значительную избыточность (соответствующие значения разборчивости 84.4 против 98.9 и 96.7%). Значение разборчивости для таблиц парных звуко сочетаний оказалось ниже, чем для таблиц выбора слов из-за отсутствия в звуко сочетаниях лексического значения (по мнению Л.Р. Зиндера [9] звуко сочетания могут иметь некоторое условное значение). Это не вызывает удивления, так как принцип дихотомии, столь удобный для описания системы различительных признаков фонем в пространстве акустико-артикуляторных коррелят и использованный при составлении диагностических таблиц, является упрощенным. В реальности взаимосвязь артикуляторных и перцептивных признаков много сложнее. Покажем это на примере применения метода диагностических таблиц для измерения свойств дифференциального признака звонкий (глухой). С этой целью была составлена испытательная таблица, отрывок из которой приведен ниже:

Быс-пыс	зэс-сэс	фиж-фиш	дац-тац	здам-стам	маж-маш
Дым-тым	крыз-крыс	нюдь-нють	зой-сой	даз-дас	зуг-зук
Гуть-куть	вяв-вяф	грыс-крыс	вур-фур	руд-рут	ребь-репь
Ждось-щтось	ег-ек	суц-зуц	дамь-тамь	чтаз-чтас	скез-скес
Дрит-трит	бяз-бяс	гар-кар	жась-шась	лув-луф	луд-лут

Таблица 2

Согласные	Вероятность встречаемости в русской речи, %		Число (%) согласных в испытательной таблице	
	твердых	мягких	твердых	мягких
/ж, ш/	4.36	–	40 (5.76)	–
/б, п/	6.32	1.52	48 (6.90)	12 (1.74)
/г, к/	6.71	0.91	48 (6.90)	12 (1.74)
/в, ф/	6.89	0.72	48 (6.90)	12 (1.74)
/з, с/	8.41	2.56	48 (6.90)	12 (1.74)
/д, т/	12.71	5.47	84 (12.06)	36 (5.22)
Итого	45.40	12.18	316 (45.42)	84 (12.18)

Таблица 3

Согласные	Число ошибок, % в зависимости от вида сигнала возбуждения		
	Импульсы ОТ	Импульсы с частотой 140 Гц	Шум
Звонкие	3.0	6.9	2.5
Глухие	5.5	8.1	8.7
В том числе:			
Начальные			
звонкие	1.6	3.1	0.7
глухие	0.9	1.6	3.1
Конечные			
звонкие	1.4	3.8	1.8
глухие	4.6	6.5	5.6

Полная таблица состояла из 400 пар слогов, в половине которых в случайном порядке менялся начальный согласный, а в другой половине – конечный согласный. Произношение в звукосочетаниях СГС конечного звонкого без оглушения требовало от диктора в данном эксперименте некоторого внимания. В табл. 2 приведен звуковой состав русской речи по согласным и состав испытательной таблицы по начальным согласным.

Такой же состав имеют конечные согласные. Звонкие и глухие согласные включены в таблицу в одинаковом количестве.

Испытания проводились с помощью установки, включавшей в себя спектрально-полосный вокодер. Вокодер содержал 19 каналов равноартикуляционной ширины в полосе частот 150–7000 Гц. Для выделения из речи сигнала основного тона ОТ использовался ларингофон и пиковый выделитель импульсов ОТ. Сигнал тон-шум формировался из импульсов ОТ. Наличие импульсов соответствовал сигнал тон, а их отсутствию – шум. Вокодер с

указанными устройствами выделения сигналов ОТ и тон-шум обеспечивал высокие качественные показатели: слоговая разборчивость речи 94–96%, а качество синтезированной речи оказалось близким к качеству передачи в стандартном телефонном канале.

Испытательная таблица была записана для трех режимов. В режиме 1 сигнал возбуждения синтезатора формировался по сигналу тон-шум из импульсов ОТ и сигнала генератора шума, в режиме 2 на синтезатор поступали импульсы с частотой 140 Гц, в режиме 3 – шумовой сигнал.

При записи диктор читал один из двух слогов каждой пары (в испытательной таблице этот слог был подчеркнут). Задача слушателей состояла в том, чтобы отметить принятый слог по отпечатанной испытательной таблице. Далее подсчитывалось число ошибок по начальным и конечным согласным, характеризующее влияние вида сигнала возбуждения на различие согласных по признаку звонкий/глухой. В опытах участвовали два диктора (мужчина и женщина) и четыре слушателя. Результаты измерения различимости согласных по признаку звонкий/глухой приведены в табл. 3.

Из приведенных данных видно, что число ошибок по глухим согласным больше, чем по звонким при всех видах сигнала возбуждения. Однако заметного влияния вида сигнала возбуждения на различимость синтезированных согласных по признаку звонкий/глухой не наблюдалось. Напротив, звонкие согласные при шумовом возбуждении различались даже лучше, чем при тональном монотонном (соответствующее число ошибок 2.5 и 6.9%).

Можно также отметить, что число ошибок во всех трех режимах различается незначительно (8.5; 15 и 11.2%). Поэтому разборчивость синтезированной речи в указанных режимах также оказывается близкой. Измерения с участием бригады из четырех дикторов и четырех слушателей показали, что слоговая разборчивость речи в режимах 1, 2 и 3 равна 93.9, 94.0 и 94.3% соответственно.

Сравним спектры звонких согласных и соответствующих глухих. На частотах выше 1000 Гц спектры указанных звуков различаются незначительно. В области частот ниже 600 Гц спектральная плотность у звонких согласных выше, чем у глухих на 20–40 дБ. Это различие формы спектров сохраняется в синтезированных звуках независимо от вида сигнала возбуждения – тонального или шумового, что и способствует правильно-му восприятию. С другой стороны, ограничение полосы речевого сигнала в области низких частот уменьшает различие между формами спектров согласных, что должно ухудшить их восприятие. В табл. 4 приведены данные, полученные в ре-

Таблица 4

Режим	Вид сигнала возбуждения	Число ошибок (%) при ограничении полосы частот, Гц	
		70–7000	600–7000
1	Импульсы ОТ	2.8/4.9*	3.4/5.5
2	Импульсы частотой 140 Гц	4.5/10.0	22.0/20.0
3	Шум	3.5/6.2	10.0/12.0

* В числителе дроби – число ошибок для звонких согласных, в знаменателе – для глухих.

Таблица 5

Режим	Вид сигнала возбуждения	Общее качество звучания, балл	Число ошибок, %	
			Узнаваемость диктора	Восприятие интонации голоса
1	Импульсы ОТ	7.8	11	3 (2/3)*
2	Импульсы частотой 140 Гц	4.2	25	49 (52/45)
3	Шум	2.9	17	24 (40/7)

* В числителе дроби указано число ошибок при восприятии интонации мужского голоса, в знаменателе – женского.

результате прослушивания таблиц при включении в тракт фильтра верхних частот с граничной частотой 600 Гц.

Из таблицы видно, что ограничение полосы частот речевого сигнала почти не увеличило числа ошибок в режиме 1, так как в этом режиме сохраняется признак вида спектра – тон или шум. Вместе с тем, в режиме 2 и особенно в режиме 3 введение ограничения полосы низких частот резко увеличивает число ошибок, хотя некоторые другие признаки различения согласных сохраняются. Действительно, звонкие и глухие коррелятивных пар фонем, например /в-ф; з-с/, кроме вида спектра различаются также формой спектра (звонкие являются более низкими), длительностью (звонкие короче глухих, в том числе фаза смычки также меньше), интенсивностью (звонкие громче), частотой основного тона на переходе от согласного к гласному [3, 10].

Рассмотрим влияние вида сигнала возбуждения на качество синтезированной речи. В режимах 1–3 были сделаны записи стандартных фраз, прочитанных шестью дикторами (тремя мужчинами и тремя женщинами). Оценка качества речи по десятибалльной шкале проводилась методом парных сравнений с участием шести – десяти слушателей. Для сравнения использовалась запись фраз естественной речи в полосе частот 300–3400 Гц.

Полученные результаты приведены в табл. 5.

Видно, что общее качество звучания синтезированной речи в большой степени зависит от вида сигнала возбуждения. Наименьшее число ошибок

было получено при возбуждении синтезатора импульсами основного тона.

Кроме общей оценки качества звучания синтезированной речи были получены также оценки по узнаваемости дикторов и восприятию интонации.

Эксперимент по узнаваемости дикторов проводился следующим образом. Были сделаны записи шести стандартных фраз, прочитанных четырьмя дикторами-мужчинами в случайном порядке. В конце каждой фразы диктор называл свой условный номер. Слушатели были незнакомы с голосами дикторов. Для знакомства перед началом испытаний все дикторы дважды прочитали по очереди одну фразу. В результате прослушивания всей записи каждый слушатель давал 24 оценки. Как и следовало ожидать, наиболее высокая узнаваемость получена в режиме 1, хотя узнаваемость дикторов в режимах 2 и 3 оставалась довольно высокой.

Для оценки влияния вида сигнала возбуждения на восприятие интонации была использована следующая методика. В режимах 1–3 были сделаны записи стандартной фразы, прочитанной диктором с тремя видами интонациями: вопрос, утверждение, приказ. Слушатели должны были дать ответ: в какой интонации прочитана фраза. В испытании участвовали восемь дикторов (пять мужчин и три женщины) и десять слушателей. Полученные результаты приведены в табл. 5, из которой видно, что в режиме 1 интонация передавалась хорошо, а в режимах 2 и 3 – с сильными искажениями. Обращает на себя внимание тот факт, что при шумовом возбуждении число ошибок на женском голосе оказалось небольшим

Таблица 6

S	35	40	45	50	55	60	65	70
Q	61	65	69	73	77	80	83	86
S	75	80	85	87.5	90	92.5	95	97.5
Q	89	92	95	96	97	98	99	99.5

(всего 7%), в то время как на мужских достигло 40%. Возможно этому способствовало высокое значение основного тона на женских голосах, при котором разрешающая способность синтезатора по частоте оказалась достаточной.

Приведенные данные достаточно убедительно свидетельствуют в пользу утверждения, что перцептивные параметры, определяющие фонетическое качество звучания – тембр, громкость, высота голоса – соотносятся не с отдельными акустическими параметрами речи, а с их совокупностью – спектром, интенсивностью, частотой основного

тона, длительностью ударных и безударных слогов. Признак звонкий представлен в звучащей речи не только гармониками основного тона, но и увеличенной интенсивностью звонких сегментов, более четкой формантной структурой по сравнению с глухими сегментами: признак мягкий отображен в *i*-образном треке второй форманты F2 на смежном гласном, частотой F2 около 1.8–2.6 кГц, а также сниженной интенсивностью составляющих в полосе частот 1.2–1.8 кГц. Следствием сказанного является слабая чувствительность таблиц DRT к искажениям отдельных признаков речевого сигнала.

Метод дописывания. В стандарт ГОСТ Р 50840-95 включены диагностические таблицы по методу дописывания начального/конечного согласного в неполных ГС/СГ слогах [3]. В качестве таблиц диктора используются стандартные слоговые таблицы. Пример таблицы диктора и таблицы слушателя (отрывки) приведен ниже.

Таблица диктора:

СЯК	КУЛЬ	СЮЩ	НЯТ	СВУМ
КОСЬ	ПИШ	СТАЛ	МУРЬ	ХАК

Таблица слушателя:

<u>ЯК</u> _АК	<u>ЮЛЬ</u> _УЛЬ	<u>ЮЩ</u> _УЩ	<u>ЯТ</u> _АТ	<u>ВУМ</u>
<u>ЕСЬ</u> _ОСЬ	ПИ_	<u>ТАЛ</u>	<u>ЮРЬ</u> _УРЬ	ХА_

В таблицах слушателя отсутствуют начальные/конечные согласные слогов (твердые или мягкие). В процессе измерений слушатель заполняет пропуски услышанным звуком. В табл. 6 приведена зависимость между слоговой разборчивостью *S* и различимостью согласных по методу дописывания *Q*.

Метод дописывания прошел длительную экспериментальную проверку в ряде заинтересованных организаций и включен в ГОСТ Р 50840-95 в качестве метода диагностической оценки искажений речевого сигнала в каналах речевой коммуникации.

По сравнению со стандартной методикой измерений слоговой разборчивости метод дописывания значительно снижает требования к тренировке артикуляционной бригады. Были проведены измерения роста тренированности неопытной бригады в составе трех слушателей при приеме слоговых таблиц синтезированной речи по ГОСТ Р и по методу дописывания. Испытания проходили в течение трех дней при ежедневном приеме 20 таблиц, записанных через тракт синтезированной речи (вокодер LPC со скоростью передачи 4.8 кбит/с), на голосе одного диктора – мужчины. Оказалось, что коэффициент экспериментальной практики бригады Z ($Z = Aэ/Aо$, где $Aэ$ – эксперименталь-

ное значение формантной разборчивости для неопытной бригады, $Aо$ – то же для хорошо тренированной бригады) при приеме слоговых таблиц за три дня медленно увеличивался (до 0.5), а при приеме таблиц по методу дописывания превысил 0.78 на второй день. Практически, примерно за неделю ежедневной работы нетренированная бригада может достичь необходимого значения Z (более 0.9). При использовании автоматического формирования состава разовых испытательных таблиц (как это предусмотрено ГОСТ Р 50840-95) запоминаемость таблиц по методу дописывания исключена.

Рабочее место оператора ЭВМ – слушателя оборудуется ПЭВМ, звуковой картой и пакетом прикладных программ КРЕС для IBM PC. Все процедуры проведения испытаний и обработка результатов измерений автоматизированы. На экране монитора высвечиваются неполные слоговые таблицы, дописывание которых выполняется через клавиатуру. Темп представления слогов равен 3 с/слог, скорость ввода равна 20 знакам/мин и не требует специальной тренировки операторов. В результате обработки данных измерений подсчитывается различимость каждой фонемы в отдельности, групповая разборчивость коррелятивных рядов фонем и звуковая разборчивость речи для некоторой заданной статистики

встречаемости звуков в звучащей речи. В таблицах дописывания вероятность встречаемости фонем может отличаться от реальной для звучащей речи, что позволяет обеспечить достаточно высокую надежность оценки значения звуковой разборчивости (распознавания) как часто например, /т, п/, так и редко встречающихся в звучащей речи звуков, например /ФЬ/ и /ХЬ/. Таким образом, известная многолетняя дискуссия по поводу допустимого отклонения звукового состава артикуляционных таблиц от состава звучащей речи, подбора представительного корпуса текстов для оценки состава, методики фонетической транскрипции текстов и статистической обработки данных теряет свою актуальность, так как результатом измерений является оценка вероятных ошибок восприятия отдельных звуков g_i , а звуковой состав речи (например, для местных или профессиональных условий) учитывается при средневзвешенной звуковой разборчивости

$D_j = 1 - Q_j$, где $Q_j = \sum p_{ij}q_i$ – потери звуковой разборчивости, p_{ij} – вероятность встречаемости i -го звука в таблице, по j -ой статистике (например, для телефонных сетей общего пользования, военной связи и т.д.), q_i – вероятность ошибки в восприятии. Слоговая разборчивость по j – статистике встречаемости звуков в звучащей речи

$$S = 0.8D_j^3 + 0.2D_j^4.$$

В методе дописывания устранен недостаток диагностических таблиц DRT, связанный с их привязкой к модели дифференциальных признаков Р. Якобсона и обусловленной этим низкой чувствительностью к искажением речевого сигнала.

Приведем пример исследования матриц звуковых ошибок (“ослышек”), полученных в результате измерений звуковой разборчивости [11]. В измерениях разборчивости участвовала бригада из четырех дикторов (двух мужчин и двух женщин) и четырех аудиторов. Были получены полные матрицы ошибок из 36 согласных для 4-х видов сигналов, включая естественную речь (при соотношении сигнал/шум 10 дБ), две модели анализа-синтеза речи (LPC-вокодер со скоростью передачи 2.4 кб/с, полувокодер со скоростью передачи 9.6 кб/с). Объем измерений составил около 100 тыс. звуков, что обеспечило статистически устойчивые результаты (видимо впервые для русского языка)*.

* Отметим, что одно из первых исследований матриц звуковых ошибок выполнили О. Miller, Р. Nycely (1955) и (1968) [12]. В первой работе использовались 200 открытых слогов СГ, включавших 16 согласных и гласный /а/ английского языка, а во второй – минимальные пары из односложных слов. Однако ограниченный объем данных не обеспечил проведение представительного статистического анализа.

Обработка матриц ошибок выполнялась на ПЭВМ IBM PC/AT с использованием методов многомерного статистического анализа данных. Методом статистического регрессионного анализа матриц звуковых ошибок (“ослышек”) были установлены следующие ранги и веса (отн. ед.) групповых признаков фонем (в порядке снижения): твердый/мягкий – 5, звонкий/глухой – 3, сонорный/шумный – 2, смычный/щелевой – 1.

Групповые признаки, в отличие от дифференциальных, имеют достаточно устойчивые акустические характеристики независимо от стиля произношения, позиционных и комбинаторных изменений фонем. С другой стороны признаки места образования в фазах экскурсии и рекурсии оказываются непостоянными и их акустические корреляты в фазе выдержки не всегда фиксируются из-за явлений редукции и коартикуляции.

Распределение звуковых ошибок (“ослышек”) зависит от вероятности встречаемости согласных (звукотипов) в русской речи: чем частотнее звукотип, тем лучше его различимость. Различие между естественной и синтезированной речью проявляется прежде всего в увеличении коэффициента регрессии и коэффициента множественной корреляции в естественной речи. Можно предположить, что в условиях более затрудненного восприятия синтезированной речи по сравнению с естественной по различительным признакам звукотипов слушатели в большей степени ориентируются на их вероятность встречаемости.

Перцептивные звуковые ошибки (“ослышки”) обладают свойством сохранения принадлежности к групповым различительным признакам произнесенных звуков. В нашем эксперименте из общего числа ошибок в распознавании твердых согласных естественной речи, равного 8%, сохранили принадлежность к группе твердых 7% “ослышек” и лишь 1% “ослышек” отошел к группе мягких, а из общего числа “ослышек” мягких, равного 8%, сохранили групповую принадлежность 6% и перешли в группу твердых – 2%. Таким образом, по признаку твердый/мягкий групповую принадлежность сохранили 75...88% “ослышек”.

Такой же характер имеют данные по различимости других признаков: по групповому признаку звонкий/глухой из общего числа ошибок, равного 5–12%, групповую принадлежность сохранили 75–80% “ослышек”, по признаку сонорный из общего числа ошибок, равного 4–80%, по признаку смычный из 4% ошибок – 70% и по признаку щелевой из общего числа ошибок 13–75%. Интересно отметить, что из всех “ослышек”, потерявших признак сонорный, 2/3 перешли в смычные и 1/3 в щелевые, а “ослышки”, потерявшие признаки щелевой и смычный, получили поровну признаки

сонорный и смычный вместо первого, сонорный и щелевой вместо второго.

В синтезированной речи групповые различительные признаки звукотипов менее устойчивы по сравнению с естественной речью. При увеличении ошибок в распознавании согласных по всем группам примерно в 2–2.5 раза относительное число “ослышек, потерявших групповую принадлежность различительных признаков произнесенных звуков, возросло по признакам: сонорный – до 50% (вместо 20% для естественной речи), в том числе получивших признак смычный – 15% и щелевой – 35%; смычный – 70% (вместо 30% для естественной речи), в том числе получивших признак сонорный – 18% и щелевой – 2% и, наконец, щелевой – 48% (вместо 25% для естественной речи), в том числе получивших признак сонорный – 16% и смычный – 32%. По различительным признакам твердый/мягкий и звонкий/глухой соотношение числа “ослышек”, сохранивших групповую принадлежность исходным звукам или сменивших ее, для естественной и синтезированной речи изменилось мало. Исходя из сказанного, можно предложить такое объяснение определенной неадекватности показателей разборчивости синтезированной речи и абонентской оценки внятности (понятности) фраз: внятность определяется не только средней различимостью согласных (в нашем примере по признаку твердый/мягкий – 82–84%, звонкий/глухой – 78–88, сонорный – 89, смычный – 85 и щелевой – 77%), но также устойчивостью сохранения частных различительных признаков звукотипов искаженной речи.

СПИСОК ЛИТЕРАТУРЫ

1. Покровский Н.Б. Расчет и измерение разборчивости речи. М.: Связьиздат, 1962. 392 с.
2. Михайлов В.Г., Златоустова Л.В. Измерение параметров речи. М.: Радио и связь, 1987. 168 с.
3. Mikhailov V.G. Speech quality measurements // Proc. of Int. workshop “Specom-99”. 4–7 oct. 1999, Moscow. P. 149–152.
4. Voiers W.D. Diagnostic Acceptability Measure for Speech Communication Systems // Proc. 1977. IEEE ICASSP, P. 204–207.
5. Вемян Г.В. Передача речи по сетям электросвязи. М.: Радио и связь, 1985. 272 с.
6. ГОСТ Р 50840-95. Передача речи по трактам связи. Методы оценки качества, разборчивости и узнаваемости.
7. Речевые тесты и их применение / Под ред. Л.В. Златоустовой. М.: МГУ, 1986. 100 с.
8. Потапова Р.К. Речь: Коммуникация, информатика, кибернетика. М.: Радио и связь, 1997. 493 с.
9. Зиндер Л.П. Общая фонетика. М.: Высшая школа, 1979. 312 с.
10. Зубкова Л.Г. Фонетическая реализация консонантных противопоставлений в русском языке. М.: УДН, 1974. 116 с.
11. Михайлов В.Г. Экспериментальная интерпретация фонологической системы русского языка (перцептивный аспект) // Акуст. журн. 1993. Т. 39. № 5. С. 879–887.
12. Miller G., Nicely P. An Analysis of Perceptual Confusion among Consonants // J. Acoust. Soc. Amer., 1995. V. 27. № 2. P. 338–352; Singh S. A Distinctive Feature Analysis of Responses to a Multiple Choice Intelligibility Test // IRAL, 1968. № 6. P. 37–53.

Diagnostic Articulation Tables

V. G. Mikhailov

Faculty of Philology, Moscow State University, Vorob'evy gory, Moscow, 119899 Russia
e-mail: rfc-lsve@mtu-net.ru

Abstract—In recent years, considerable progress has been made in the development of instrumental methods for the general speech quality and intelligibility evaluation on the basis of modeling the auditory perception of speech and measuring the signal-to-noise ratio. Despite certain advantages (fast measurement procedures with a low labor consumption), these methods are not universal and, in essence, secondary, because they rely on the calibration based on subjective-statistical measurements. At the same time, some specific problems of speech quality evaluation, such as the diagnostics of the factors responsible for the deviation of the speech quality from standard (e.g., accent features of a speaker or individual voice distortions), can be solved by psycholinguistic methods. This paper considers different kinds of diagnostic articulation tables: tables of minimal pairs of monosyllabic words (DRT) based on the Jacobson differential features, tables consisting of multisyllabic quartets of Russian words (the choice method), and tables of incomplete monosyllables of the *_VC/CV_* type (the supplementary note method). Comparative estimates of the tables are presented along with the recommendations concerning their application.