Речевые технологии

1/2009

Содержание

ИНФОРМАЦИОННЫЕ РЕСУРСЫ

КОНФЕРЕНЦИЯ КОМПЬЮТЕРНЫЕ РЕЧЕВЫЕ ТЕХНОЛОГИИ

Редакция:
Редактор — *Артём Ганькин*
Корректор — *Татьяна Денисьева*
Дизайн — *Анна Ладанюк*
Вёрстка — *Сергей Бурукин*

Адрес редакции: 109341, Москва, ул. Люблинская, д. 157, корп. 2.
Тел.: 8 (495) 979-54-27

Подписано в печать 24.09.2009. Формат 60×90⅛. Бумага офсетная. Печать офсетная.
Печ. л. 6. Заказ № 1002. Издательский дом «Народное образование».
Отпечатано в типографии НИИ школьных технологий. 143500, г. Истра-2, ул. Заводская, д. 2А.
Тел.: 8 (901) 513-97-64, (495) 792-59-62.

Устойчивость оценок формантных частот

В.Н. Сорокин
доктор физико-математических наук

А.С. Леонов,
доктор физико-математических наук

И.С. Макаров

Выполнен сравнительный анализ точности и устойчивости мгновенных оценок формантных частот в речевом сегменте методом нулей сигнала и различными модификациями метода линейного предсказания для синтезированных звуков и сигналов, параллельно записанных с микрофонов разного типа. Все использованные методы линейного предсказания показали существенно больший разброс оценок, чем метод нулей сигнала. Установлено, что стабилизация мгновенных оценок формантных частот достигается путём использования информации о характерных акустических характеристиках гласноподобных звуков в конкретном языке. Устойчивость определения формантных треков обеспечивается путём их аппроксимации кусочно-линейными функциями.

Введение

Для решения обратной задачи нахождения формы речевого тракта по сегменту речи нужно оценить резонансные частоты тракта, используя речевой сигнал. Он, однако, определяется не только резонансами речевого тракта, но и резонансами подсвязочной области — трахеи, бронхов и лёгких. Кроме того, в нём присутствуют резонансы носовой полости, причём не только для назальных согласных, но и для назализованных гласных. Поэтому выбор резонансных частот, принадлежащих только речевому тракту выше голосовой щели, представляет собой трудную задачу. Более того, оценка формантных частот тракта по сигналу есть некорректно поставленная задача, что может выражаться в неоднозначности решения (при наличии близких резонансов) и его неустойчивости по отношению к погрешностям измерений. Последние связаны с искажениями сигнала каналом регистрации, реверберацией помещения, нестабильностью частоты основного тона и другими факторами. Амплитудные и частотные модуляции формант усугубляют неоднозначность оценок их частоты.

Оценки формантных частот выполняют как в частотной, так и во временной области. Один из самых распространённых подходов основан на ***методах линейного предсказания***, которые предназначены для описания сигнала во временной области. В целом, эти методы могут давать удовлетворительные оценки формант. Однако многолетние исследования такого подхода показали, что любые модификации методов линейного предсказания неустойчивы относительно аддитивных шумов, особенно при оценке низкочастотных формант. Даже при относительно хороших условиях измерений погрешность оценки формант методами линейного предсказания, как правило, не ниже 10% и к тому же зависит от частоты основного тона [1].

Метод нулей сигнала для оценки формантных частот [5, 6] основан на анализе распределения длительностей интервалов между нулями сигнала. Идеи, лежащие в основе метода, использованы ещё в первых работах по исследованию проблемы автоматического распознавания речи. В своих первых реализациях (на аналоговых устройствах) метод обычно применялся к так называемому клиппированному сигналу. Последний получался путём использования очень большого коэффициента усиления с последующим ограничением амплитуды. В результате преобразованный речевой сигнал представлялся в виде последовательности прямоугольных импульсов с фиксированной амплитудой [2]. Это было удобно для обработки сигнала аналоговой аппаратурой. Но оказалось, что клиппированный сигнал имеет низкую помехоустойчивость, и в результате метод оценки формант с помощью выделения нулей сигнала был на какое-то время забыт.

Развитие цифровой техники привело к возрождению интереса к методу нулей сигнала. В своём новом варианте [5,6] метод оказался более помехоустойчивым, чем методы линейного предсказания и спектрального анализа. Кроме того, метод нулей позволяет обнаружить тонкую структуру динамики формант [5,6]. В работах [3,4] показано, что один из вариантов метода нулей, рассмотренный там под названием «zero-crossing», превосходит известные методы линейного предсказания для низких формант вплоть до SNR=0 dB.

1. Алгоритм метода нулей сигнала

Особенность метода нулей заключается в игнорировании формы колебаний. При этом, конечно, теряется часть информации. Поэтому при низком уровне шумов такие методы, как автокорреляционный или линейное предсказание, могут иметь преимущество. Однако форма колебаний искажается по мере роста уровня шумов, и это преимущество превращается в недостаток.

В большинстве методов оценки формант применяется предварительная обработка сигнала с помощью набора пересекающихся полосовых фильтров. Тип фильтров, их полоса и степень перекрытия влияют на качество последующего анализа и итоговых оценок резонансных частот. Использование фильтров в полосах, примерно соответствующих диапазонам положения формант, способствует повышению точности и устойчивости оценок. После подобной предварительной обработки иногда применяется адаптивный фильтр для уточнения положения формант формантных частот [7]. Анализ нулей сигнала предполагает, что в данной частотной полосе присутствуют колебания только одной форманты. Это связано с известным свойством,

согласно которому при наличии нескольких частот средняя частота переходов определяется как средневзвешенная по амплитудам каждой частоты. Именно поэтому в методе нулей сигнала особенно важен выбор полос частот для анализа.

В данной работе рассматриваются три метода предварительной фильтрации сигнала в частотных диапазонах, где ожидается присутствие только одной форманты.

В первом методе частотные полосы фильтров устанавливаются следующим образом. Первая форманта любого звука анализируется в двух фильтрах с полосами 130 Гц — 400 Гц (фильтр $Ф_{11}$) и 300 Гц — 800 Гц (фильтр $Ф_{12}$). Второй форманте соответствуют три фильтра: 700 Гц — 1600 Гц (фильтр $Ф_{21}$), 1000 Гц — 2000 Гц (фильтр $Ф_{22}$) и 1400 Гц — 2400 Гц (фильтр $Ф_{23}$). Наконец, третья форманта ожидается в одном из двух фильтров с полосами 1700 Гц — 2500 Гц (фильтр $Ф_{31}$) и 2300 Гц — 3500 Гц (фильтр $Ф_{32}$). Эти фильтры перекрываются, в результате чего в один фильтр могут попасть колебания, отвечающие двум формантам.

Второй метод использует распределения формант для каждого гласного русского языка в предположении, что тип гласного и пол диктора известны. Диапазоны формантных частот для некоторых русских гласных даны в таблицах 1, 2.

Таблица 1

Диапазоны формантных частот гласных русского языка для мужских голосов

Гласный	F1 Гц	F2 Гц	F3 Гц
А	450–850	950–1500	1900–2950
Э	320–530	1450–2250	2000–2950
О	300–750	600–1400	1800–3200
И	200–550	1650–2750	2250–3500
Ы	210–500	1650–2600	2150–3100
Е*	250–570	1450–2550	2150–3350
Я*	330–750	1350–2200	2000–3100

* В позиции между мягкими согласными.

Таблица 2

Диапазоны формантных частот гласных русского языка для женских голосов

Гласный	F1 Гц	F2 Гц	F3 Гц
А	550–1000	1100–1650	1950–3100
Э	350–600	1800–2600	2350–3350
О	320–850	600–1550	1800–3300
И	220–620	1850–3100	2550–3600
Ы	250–580	1900–2950	2300–3600
Е*	300–650	2000–2950	2650–3650
Я*	400–900	1800–2650	2300–3500

* В позиции между мягкими согласными.

Третий метод использует параллельный анализ сигнала в полосах, характерных для всех гласных русского языка, в случае когда тип гласного неизвестен или наблюдается переход от одного гласного к другому. Окончательный выбор оценок формантных частот выполняется по критерию, включающему вероятность попадания в трёхмерный вектор формантных частот и суммарную энергию сигнала на этих частотах. Если неизвестен и пол диктора, то сигнал анализируется в частотных полосах, установленных и для мужчин, и для женщин. При этом может использоваться информация, найденная из анализа формы голосового источника. Как показано в [8], вероятность правильного определения пола диктора составляет около 90%. С теоретической точки зрения, частоты резонансов речевого тракта могут изменяться на периоде основного тона вследствие изменения граничных условий при переходе от открытой голосовой щели к закрытой. Кроме того, частоты формант в речевом сигнале подвержены влиянию голосового источника. Поэтому оценку формантных частот целесообразно выполнять на интервале закрытой голосовой щели. В данной работе этот интервал определяется как область, примерно равная 30% от периода основного тона, смещённая на 1 мс относительно пиков огибающей по Гильберту в каждой частотной полосе. Эти пики соответствуют всплеску энергии колебаний резонанса после смыкания голосовой щели.

Во всех методах после фильтрации исходного сигнала с помощью каждого из используемых фильтров определяется среднее значение разности времени между нулями отфильтрованного сигнала на интервале закрытой голосовой щели. Это значение принимается как оценка полупериода формантного колебания из рассматриваемого частотного диапазона. Если оказывается, что нулей меньше двух, то оценка не производится. Затем находится среднее значение формантной частоты для нескольких периодов основного тона, формируется узкополосный фильтр с центральной частотой, равной этой средней частоте, и после новой фильтрации исходного сигнала уточняются оценки частот колебаний на данном интервале времени.

На этом же интервале времени вычисляется среднее значение энергии колебаний, и в качестве предварительной оценки частоты форманты выбирается оценка из того фильтра, где энергия наибольшая. При этом отсеиваются оценки, выходящие за пределы диапазона, а среди конкурирующих оценок выбирается та, которая ближе к среднему значению диапазона.

2. Сравнительное тестирование методов формантного анализа

Ниже приводятся результаты сравнения оценок формант методами типа линейного предсказания и методом нулей сигнала. Изучалась точность и устойчивость методов по отношению к аддитивным помехам, типу микрофона и реверберации помещения, а также устойчивость оценки формантных частот в естественной речи.

2.1. Устойчивость относительно аддитивного белого шума

Погрешность определения формантных частот обычно оценивается в экспериментах с синтезированными звуками, поскольку нет другого способа полу-

чить сигнал с известными параметрами. Однако этому методу присущи недостатки, которые не позволяют безоговорочно опираться на результаты такого тестирования. Синтетический сигнал — это суперпозиция колебаний нескольких осцилляторов с собственными частотами, близкими к реальным формантным частотам, под воздействием источника возбуждения, который по своим характеристикам близок к реальному голосовому источнику. Возбуждаемые этим источником парциальные колебания отличаются от собственных колебаний осцилляторов даже на временных участках, соответствующих закрытой голосовой щели. Это отличие вносит ошибку в оценки собственных частот.

Сигнал, синтезированный с помощью суммирования экспоненциально затухающих колебаний, должен был бы наилучшим образом соответствовать анализу методом линейного предсказания, где используется модель, состоящая из набора полюсов. Поэтому этот метод имеет преимущество перед методами, не опирающимися на такую модель. Тем не менее, в присутствии помех даже для синтезированных сигналов метод линейного предсказания не всегда оказывается наилучшим.

Эксперименты по сравнительной оценке точности и устойчивости методов анализа частотного состава синтетических гласных */А, И, У/* проводились при наличии помех типа белого шума разного уровня с гауссовым распределением. Для каждого уровня шумов проводилось по 100 испытаний. Использовались разные варианты линейного предсказания: автокорреляционный, ковариационный, метод усечённого сингулярного разложения матриц, метод регуляризации по Тихонову и метод DAP, разработанный в [9] специально для повышения точности анализа женских голосов.

Оценки формантных частот, полученные с помощью линейного предсказания в кратковременном окне анализа, подвергались сортировке. Из множества исходных оценок удалялись действительные полюсы, а также комплексно-сопряжённые полюсы, частота которых ниже некоторого порога (например, 200 Гц). Кроме того, удалялись полюсы, ширина которых превышает некоторый порог (например, 500 Гц).

Результаты анализа этими методами и разработанным нами методом нулей сигнала показаны в таблицах 3, 4 и 5.

Таблица 3

Относительные ошибки (в %) вычисления формантных частот при отношении сигнал/шум SNR = 20 dB

	Autocorrelation LPC			Covariance LPC			DAP			Метод нулей		
	dF1	dF2	dF3	dF1	dF2	dF3	dF1	dF2	dF3	dF1	dF2	dF3
A	-0.8	-0.2	-0.6	-0.3	-0.3	-0.4	-0.1	-0.1	-0.7	-3.6	-1.2	2.0
I	5.8	-1.1	0.3	4.3	-1.8	0.3	2.5	-1.1	0.1	-5.7	-2.8	-0.4
U	11.5	7.2	-5.9	11.1	6.7	-6.0	6.4	5.6	-6.0	-11.1	2.7	2.8

Таблица 4

Относительные ошибки (в %) вычисления формантных частот при отношении сигнал/шум SNR = 15 dB

	Autocorrelation LPC			Covariance LPC			DAP			Метод нулей		
	dF1	dF2	dF3	dF1	dF2	dF3	dF1	dF2	dF3	dF1	dF2	dF3
A	-1.7	-0.5	-0.5	-0.9	-0.6	-0.6	-0.2	-0.4	-0.4	-3.7	-2.2	3.2
I	7.5	0.7	0.6	6.2	0.9	0.6	3.7	0.4	0.4	-5.8	-7.4	-0.4
U	26.9	17.7	-5.9	26.8	17.6	-5.8	18.5	13.0	-6.1	-10.9	6.0	2.4

Таблица 5

Относительные ошибки (в %) вычисления формантных частот при отношении сигнал/шум SN = 10 dB

	Autocorrelation LPC			Covariance LPC			DAP			Метод нулей		
	dF1	dF2	dF3	dF1	dF2	dF3	dF1	dF2	dF3	dF1	dF2	dF3
A	-1.6	-0.0	-1.6	-0.6	-0.0	-0.6	-2.0	-0.2	-0.1	-3.9	-1.7	4.3
I	11.2	0.9	0.8	10.2	1.0	0.8	6.2	0.7	0.7	-5.6	-13.3	3.2
U	36.2	35.0	-3.0	36.0	35.0	-3.1	31.1	28.3	-3.0	-11.1	15.8	4.8

Таблицы подтверждают установленное другими исследователями свойство неустойчивости к шумам оценок формант методами линейного предсказания, особенно заметное при оценке низких частот. Оценка методом нулей сигнала несколько уступает по точности методам линейного предсказания при низком уровне шума, но оказывается значительно устойчивее при высоком уровне шума. Этот же вывод действителен и для других испытанных нами методов линейного предсказания, не показанных в таблицах 3–5.

2.2. Устойчивость относительно типа микрофона

Сравнение точности и устойчивости методов определения формантных частот с использованием только синтетических звуков не гарантирует полной объективности. Именно поэтому мы провели сравнение результатов оценки формант одного и того же речевого сигнала, записанного одновременно разными приемниками звука. Разница в вычисленных формантах характеризует устойчивость метода относительно искажений амплитудно-частотной характеристики канала связи.

Эксперименты по сравнению устойчивости методов относительно типа микрофона выполнялись на речевых сигналах, отобранных из базы данных для русских числительных. В первой группе дикторов речевые сигналы записывались параллельно через телефонную трубку в стандартном положении и в направленный микрофон, укреплённый вертикально на груди диктора. Во второй группе дикторов использовалась телефонная трубка другого типа и кардиоидный микрофон на груди диктора. В третьей группе дикторов

речевой сигнал записывался через микрофон на головной гарнитуре и кардиоидный микрофон, установленный на мониторе компьютера на расстоянии примерно 50–70 см от диктора. Для экспериментов были случайно отобраны по одному мужчине и одной женщине из каждой группы. Из речевых сегментов каждого числительного были вырезаны стационарные участки ударных гласных, которые и подвергались анализу.

Результаты сравнения для метода нулей сигнала и метода DAP приведены в таблицах 6–9.

Таблица 6

Расхождение оценок формантных частот (%). Метод нулей сигнала. Мужчины

	Первая группа			Вторая группа			Третья группа		
Гласный	**dF1**	**dF2**	**dF3**	**dF1**	**dF2**	**dF3**	**dF1**	**dF2**	**dF3**
нОль	0.0760	0.0578	0.0972	0.0324	0.0216	0.0975	**0.2002**	0.0449	0.0273
одИн	0.0468	0.0221	0.0918	**0.1682**	0.0069	0.0081	0.0783	0.0283	0.0010
двА	0.0096	**0.1015**	0.0037	0.0246	0.0295	**0.1320**	0.0321	0.0062	0.0095
трИ	0.0644	0.0488	**0.1312**	**0.1560**	0.0479	0.0548	0.0632	0.0003	0.0953
четЫре	0.0571	0.0386	0.0595	**0.2160**	0.0723	0.0645	**0.1071**	0.0050	**0.1618**
пЯть	0.0091	0.0228	0.0912	0.0272	0.0136	0.0083	0.0011	0.0118	**0.1881**
шЭсть	0.0015	0.0120	0.0568	0.0665	0.0430	**0.1058**	**0.1046**	0.0150	0.0826
сЕмь	0.0060	0.0062	0.0826	0.0661	0.0158	0.0136	**0.2350**	0.0174	0.0556
вОсемь	0.0182	**0.1325**	0.0029	0.0888	0.0739	**0.2224**	**0.1012**	0.0372	0.0742
дЕвять	0.0085	0.0043	0.0295	**0.2536**	0.0435	**0.1017**	**0.1060**	0.0002	**0.1035**
Среднее	0.0297	0.0447	0.0646	0.1099	0.0368	0.0809	**0.1029**	0.0166	0.0799

Среднее — 6.4%

Средняя ошибка оценки формант по методу нулей сигнала для всех гласных у мужчин составляет: в первой группе — 4.6%, во второй группе — 7.6%, а в третьей группе — 6.6%. Количество рассогласований оценок с уровнем от 10% до 20% равно 14, а с уровнем от 20% до 30% равно 5.

Таблица 7

**Расхождение оценок формантных частот (%).
Метод линейного предсказания DAP с предыскажением. Мужчины**

	Первая группа			Вторая группа			Третья группа		
Гласный	**dF1**	**dF2**	**dF3**	**dF1**	**dF2**	**dF3**	**dF1**	**dF2**	**dF3**
нОль	0.0603	0.0380	**0.1248**	0.0116	0.0649	0.0011	0.0246	0.0718	0.0271
одИн	0.0323	0.0608	0.0640	**0.1633**	0.0206	0.0403	**0.2977**	0.0170	0.0312
двА	0.0009	0.0641	**0.2444**	0.0053	0.0166	0.0137	0.0482	0.0024	0.0276
трИ	**0.1020**	0.0779	0.0085	**0.2095**	**0.1604**	0.0895	0.2298	0.0385	0.0004

четЫре	0.0046	0.0302	0.0507	0.2759	0.0646	0.0170	0.1122	0.0678	0.0166
пЯть	0.0174	0.0224	0.0142	0.0395	0.0120	0.0131	0.0217	0.0080	0.0545
шЭсть	0.0012	0.0014	0.0030	**0.1184**	0.0175	0.0880	NaN	**0.1024**	0.0067
сЕмь	0.0797	0.0023	0.0270	**0.1494**	0.0308	0.0040	0.2221	0.0769	**0.1103**
вОсемь	0.0561	NaN	0.0003	0.0750	0.0269	0.0006	**0.1994**	0.0703	0.0823
дЕвять	**0.1598**	0.0126	0.0051	**0.1986**	**0.1032**	0.0079	**0.1402**	0.0166	0.0379
Среднее	0.0514	0.0344	0.0542	**0.1246**	0.0517	0.0275	**0.1440**	0.0472	0.0395

Среднее — 6.4%

Средняя ошибка оценки формант методом линейного предсказания по всем гласным у мужчин составляет: в первой группе — 4.7%, во второй группе — 7.6%, а в третьей группе — 6.6%. Так же, как и в методе нулей сигнала, количество рассогласований оценок с уровнем от 10% до 20% равно 14, а с уровнем от 20% до 30% равно 5. Без учёта грубых ошибок метода DAP средняя ошибка по всем измерениям в обоих методах одинакова. Однако имеются две грубые ошибки, когда оценка форманты по методу DAP выходит за ожидаемый диапазон значений формант.

Таблица 8

**Расхождение оценок формантных частот (%).
Метод нулей сигнала. Женщины**

	Первая группа			Вторая группа			Третья группа		
Гласный	**dF1**	**dF2**	**dF3**	**dF1**	**dF2**	**dF3**	**dF1**	**dF2**	**dF3**
нОль	0.0148	**0.1785**	0.0223	0.0654	0.0230	**0.1349**	0.0236	**0.1150**	0.0806
одИн	0.0061	0.0741	0.0691	0.0842	0.0719	0.0170	**0.1565**	0.0196	0.0492
двА	0.0224	0.0139	**0.1023**	0.0112	**0.1048**	**0.1003**	0.0185	0.0035	**0.1011**
трИ	0.0865	0.0559	0.0440	0.0276	**0.1805**	0.0026	**0.1496**	0.0379	0.0319
четЫре	**0.1445**	0.0400	0.0222	**0.1665**	0.0111	0.0057	0.1069	0.0089	0.0305
пЯть	0.0062	0.0033	0.0565	0.0304	0.0254	**0.1200**	0.0253	0.0256	0.0720
шЭсть	0.0712	0.0033	0.0152	0.0719	0.0058	0.0174	0.0246	0.0316	0.0146
сЕмь	0.0592	0.0365	**0.1065**	0.0849	0.0348	0.0238	**0.1494**	0.0214	0.0106
вОсемь	**0.1221**	0.0298	0.0636	0.0662	0.0107	0.2079	**0.1359**	**0.1606**	0.0528
дЕвять	0.0128	0.0591	0.0389	0.0394	0.0217	0.0096	0.0446	0.0792	0.0140
Среднее	0.0546	0.0494	0.0541	0.0648	0.0490	0.0639	0.0835	0.0503	0.0457

Среднее — 5.7%

У женщин средняя ошибка оценки формант по методу нулей сигнала (для всех гласных) составляет: в первой группе — 5.3%, во второй группе — 5.9%, а в третьей группе — 6.0%. Количество рассогласований оценок с уровнем от 10% до 20% равно 18, а с уровнем от 20% до 30% равно 1.

Таблица 9

**Расхождение оценок формантных частот (%).
Метод линейного предсказания DAP с предыскажением. Женщины**

	Первая группа			Вторая группа			Третья группа		
Гласный	**dF1**	**dF2**	**dF3**	**dF1**	**dF2**	**dF3**	**dF1**	**dF2**	**dF3**
нОль	0.0143	0.0304	**0.1704**	**0.2306**	0.0194	**0.1547**	**0.2103**	**0.4365**	**0.1852**
одИн	0.0066	0.0371	0.0940	**0.3084**	**0.1468**	0.0265	**0.2813**	0.0548	0.0280
двА	0.0029	0.0219	0.0389	0.0090	0.0740	0.0089	0.0011	0.0907	**0.1690**
трИ	0.0151	0.0032	0.0932	**0.4084**	**0.1343**	0.0134	**0.1176**	0.0182	0.0304
четЫре	0.0963	0.0277	0.0537	0.0959	**0.2006**	0.0334	0.0629	**0.1047**	0.0094
пЯть	**0.1451**	0.0251	0.0101	0.0232	0.0401	0.0978	**0.2363**	0.0235	0.0480
шЭсть	0.0096	0.0321	0.0326	0.2413	**0.1237**	**0.1753**	0.0155	0.0191	0.0116
сЕмь	0.0253	0.0291	0.0094	0.0954	**0.1056**	0.0784	**0.1309**	0.0072	0.0734
вОсемь	0.0664	0.0622	**0.1141**	**NaN**	**NaN**	**NaN**	**0.1155**	**0.2953**	0.0368
дЕвять	0.0223	0.0752	0.0461	**0.2611**	**0.1907**	0.0160	0.0204	0.0766	0.0324
Среднее	0.0404	0.0344	0.0663	**0.1859**	**0.1150**	0.0672	**0.1192**	**0.1127**	0.0624

Среднее — 8.9%

Средняя ошибка оценки формант методом линейного предсказания по всем гласным у женщин составляет: в первой группе — 4.7%, во второй группе — 12.3%, а в третьей группе — 9.8%. Количество рассогласований оценок с уровнем от 10% до 20% равно 15, с уровнем от 20% до 30% равно 7. Имеется одна ошибка с уровнем от 30% до 40%, и две ошибки превышают 40%. Кроме того, имеются три грубые ошибки.

Анализ выполнялся первым методом, т.е. в усреднённых диапазонах частот для каждой форманты. Однако в силу того что тип гласного известен, окончательный отбор оценок формант производился с учётом характерного диапазона формантных частот и степени близости к характерному среднему значению каждой форманты гласного.

В таблицах использован термин среды МАТЛАБ–NAN (Not a Number). Он означает, что для одного из микрофонов не найдена оценка форманты в заданном диапазоне. В силу ограниченности тестового материала, разницу в долях процентов можно считать малозначимой, тогда как разница в процентах указывает на определённую тенденцию.

При сравнении данных из таблиц 8 и 9 видно, что число грубых ошибок метода DAP, включая выход за диапазон частот и превышение ошибки в 30% равно 6. Средняя ошибка в методе DAP, специально разработанном для улучшения качества анализа женских голосов, в полтора раза больше, чем в методе нулей сигнала.

Первая и вторая группы отличаются, главным образом, вторым микрофоном, поскольку разницу между двумя типами телефонных трубок можно считать малой по сравнению с разницей между направленным и кардиоидным микрофонами. Разница в оценках формантных частот у мужчин составляет около 3% для обоих методов. Даже при ограниченном речевом материале эта разница представляется значимой. У женщин эта

разница в методе нулей сигнала составляет всего 0.6%, тогда как в методе линейного предсказания она достигает почти 8%.

Итак, оба метода чувствительны к типу микрофона, причём у женщин разница между оценками формант по сигналам от направленного и ненаправленного микрофонов особенно велика в методе линейного предсказания.

2.3. Устойчивость относительно реверберации

Акустические характеристики помещения, в котором происходит запись речевого сигнала, влияют на амплитудно-частотные характеристики сигнала. Это было наглядно продемонстрировано в [10].

Данные таблиц 6–9 позволяют качественно оценить влияние реверберации помещения на погрешность методов анализа. Первая и вторая группы тестов выполнялись на относительно близко расположенных микрофонах, тогда как в третьей группе тестов использовались и близко расположенный ко рту микрофон, и микрофон, удалённый на расстояние в несколько десятков сантиметров. При этом во второй и третьей группах тестов один из микрофонов был один и тот же — микрофон кардиоидного типа, расположенный либо на груди диктора, либо на мониторе.

Средние значения ошибок в методе нулей сигнала для близко расположенных микрофонов и удалённого микрофона оказались довольно близки: 6.1% и 6.6% — у мужчин, и 5.6% и 6.0% — у женщин, так что ошибки отличаются на величину около 0.5%. Для метода линейного предсказания эта разница оказалась больше: 5.7% и 7.5% — у мужчин, и 8.5% и 9.8% — у женщин. В этом случае различие оценок для близких и удалённых микрофонов составила 1.8% и 1.3%. Из этого можно заключить, что реверберация помещения больше сказывается на анализе методом линейного предсказания, чем на анализе методом нулей сигнала.

2.4. Устойчивость анализа натуральных звуков

Число полюсов в амплитудно-частотной характеристике речевого сигнала, оцениваемое методом линейного предсказания, связано с частотой дискретизации сигнала. Поэтому в диапазоне частот, характерных для какого-либо звука речи, может оказаться либо избыточное, либо недостаточное количество полюсов. Это вполне закономерно, поскольку метод линейного предсказания изначально предназначен для аппроксимации сигнала, а не для анализа резонансных частот речевого тракта. Поскольку коэффициенты линейного предсказания вычисляются в процедуре, которая минимизирует ошибку аппроксимации спектра, то количество найденных полюсов и их расположение, вообще говоря, произвольны. И хотя в большинстве случаев вычисленные полюса достаточно близки к резонансам речевого тракта, имеется достаточно много ситуаций, в которых появляются грубые ошибки в оценке формантных частот.

Устойчивость оценок формантных частот методом нулей сигнала зависит от параметров полосовых фильтров и от точности определения интервала

сомкнутых голосовых складок. Как следствие, метод нулей сигнала может ненадёжно определять формантные частоты при сближении формант или на переходных процессах.

Если заранее известно, какой тип гласного соответствует рассматриваемому сегменту речи, как это может иметь место при верификации диктора, то целесообразно использовать фильтры, настроенные на конкретный гласный уже на первом этапе анализа. В этом случае метод нулей сигнала демонстрирует наиболее устойчивые оценки формантных частот. Так, в описанных выше экспериментах по сравнению устойчивости оценок для сигналов, записанных параллельно с микрофонов разных типов, средняя ошибка в методе нулей сигнала для мужчин составила около 4%, а для женщин — около 3%, т.е. в 1.5–2 раза меньше, чем при анализе с помощью фильтров, настроенных на усреднённые диапазоны формант. Это сопоставимо с погрешностью оценок, возникающей из-за дискретизации сигнала по времени.

Если дополнительная информация об ожидаемом типе гласного или переходном процессе отсутствует, то ни один из известных методов анализа формантных частот, включая и разработанный нами метод нулей сигнала, не застрахован от грубых ошибок. Поэтому кажется естественным применить параллельный формантный анализ разными методами. Если бы удалось совместить сильные стороны каждого метода и избежать их недостатков путём формирования критерия выбора оценок, то можно было бы надеяться на получение более точных и устойчивых оценок формантных частот.

Один из вариантов подобного параллельного анализа состоит в использовании метода линейного предсказания для предварительной оценки формантных частот в относительно широких диапазонах возможного положения каждой форманты. Эти оценки используются затем для формирования адаптивных фильтров, выходные сигналы которых анализируются методом нулей сигнала. Недостаток такого подхода состоит в риске грубых ошибок линейного предсказания.

Поиск критерия выбора правильного решения при параллельном использовании разных методов формантного анализа требует специального исследования. В качестве альтернативы такому подходу в данной работе применялся только метод пересечений через нуль, но параллельная оценка выполнялась для всего множества фильтров, соответствующих диапазонам формантных частот каждого гласного.

3. Динамика формантных частот

Известно, что мгновенные оценки формантных частот любым методом ненадёжны. Графически это выглядит как разброс точек на формантных треках (см., например, рис. 4). При этом иногда (например, в случае близких формант) трудно определить, какому треку принадлежит какая точка. Поэтому, получив формантные треки на интервалах времени определённой длительности, обычно выполняют коррекцию ошибок кратковременного анализа путём интерполяции треков.

Многие алгоритмы коррекции формантных треков используют предположение об их непрерывности или гладкости, основанное на непрерывности артикуляторных движений. Например, в классическом алгоритме [11] сначала находятся квазистационарные согласованные сегменты речевого сигнала (так называемые опорные точки), на которых оценки формантных частот наиболее надёжны. Затем алгоритм последовательно про-

должает формантные треки между соседними опорными точками, выбирая при переходе от предыдущего к последующему формантному вектору из множества кандидатов тот, который наиболее близок (в евклидовой метрике) к уже оценённому на предыдущем сегменте. В работе [12] построение формантных треков по оценкам линейного предсказания осуществляется с помощью дискретных марковских моделей. В работах [13, 14] искомые векторы формантных частот выбираются из множества кандидатов с помощью процедуры динамического программирования. При этом используется некоторый составной критерий отбора, который включает в себя невязку соседних по времени векторов формантных частот, условие минимума формантных ширин и близость формант к формантным частотам нейтрального гласного.

Однако на участках переходных процессов в речевом сигнале нередко наблюдается нарушение непрерывности формантных треков. Это явление характерно для женских голосов, хотя у мужских голосов оно также иногда наблюдается. В качестве примера рассмотрим рис. 1 и 2, где показаны сонограммы слогов /*УА*/ и /*АУ*/ для женского голоса. Эти сонограммы демонстрируют не только разрывы треков первой и второй форманты, но и разрывы направления движения формант. Отметим, что эти разрывы не

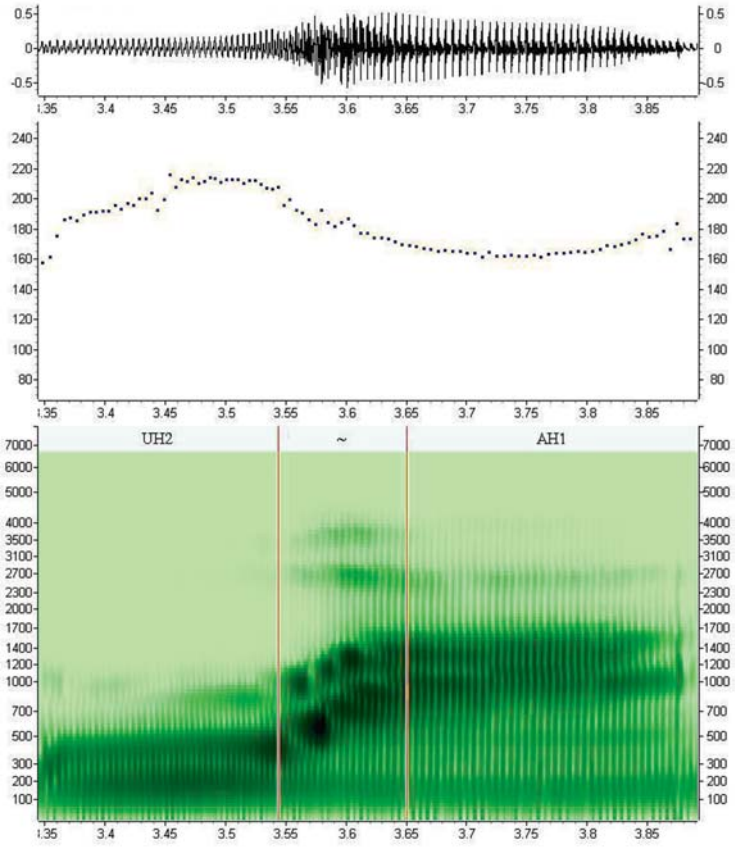


Рис. 1. Звукосочетание /УА/, женский голос.
Вверху – осциллограмма сигнала, в середине – контур основного тона, внизу – сонограмма со шкалой мел по оси частот

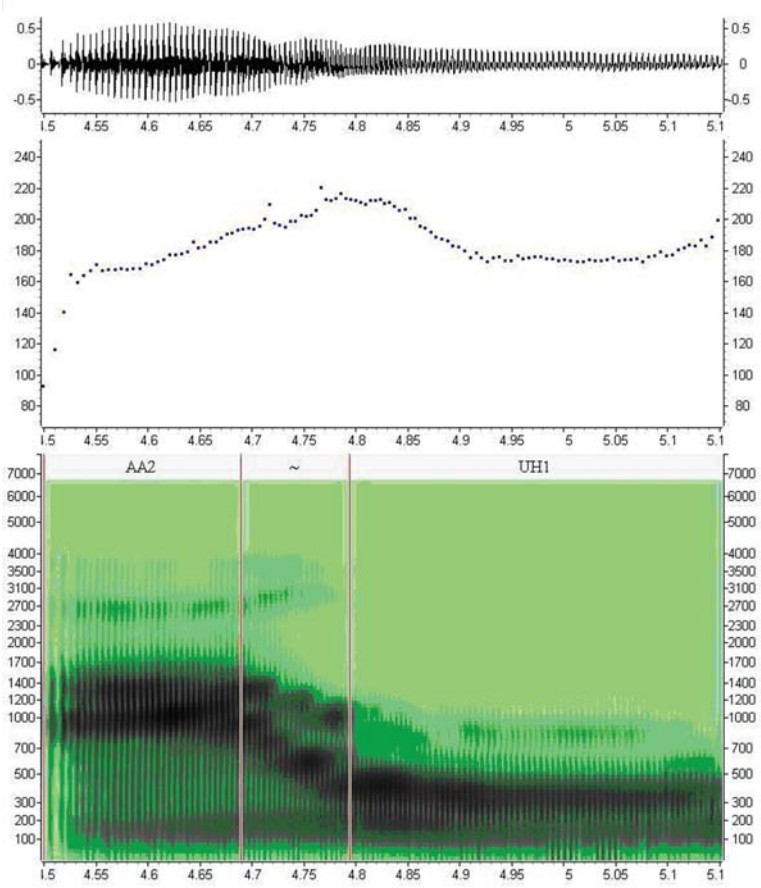


Рис. 2. Звукосочетание /АУ/, женский голос. Вверху – осциллограмма сигнала, в середине – контур основного тона, внизу – сонограмма со шкалой мел по оси частот

регистрируются методами линейного предсказания, и лишь анализ интервалов между нулями сигнала при закрытой голосовой щели обнаруживает эти явления. На обоих рисунках видно, что разрывы в формантных треках сопровождаются амплитудными модуляциями осциллограмм, и даже на квазистационарном участке трека второй форманты наблюдаются довольно длительные спады энергии. Это затрудняет использование информации об амплитуде формант при отслеживании треков оценок формантных частот во времени.

В работе [15] было показано, что разрывы формантных треков в динамическом спектре речевого сигнала могут наблюдаться в тех случаях, когда резонансные частоты речевого тракта и подсвязочной области близки. Близость ротовых и подсвязочных резонансов не является единственной причиной разрывов. Другие факторы и, в частности, соотношение частоты основного тона и частоты форманты, также влияют на форму динамического спектра звуков речи. В особенности это относится к женским голосам с высоким основным тоном. Некоторые математические аспекты этого явления рассмотрены в *Приложении*.

Из сказанного ясно, что коррекцию формантных треков необходимо выполнять, исходя из возможной их разрывности. Соответственно, при интерполяции треков нет оснований для использования непрерывных функций и, в частности, многочленов высоких поряд-

ков. Наиболее целесообразным представляется использование кусочно-линейной аппроксимации треков.

Приведём пример такой коррекции треков формант. На рис. 3 показаны мгновенные оценки формантных треков в слоге /*ИА*/ по методу нулей сигнала и их кусочно-линейная аппроксимация.

На интервале времени вокруг отсчёта 0.25 с наблюдаются скачки всех трёх формант. Особенно велик скачок частоты второй форманты (около 500 Гц). На сонограмме этого слога действительно видны разрывы траекторий формантных частот при переходном процессе от звука /*И*/ к звуку /*А*/. Однако эти скачки заглажены в силу использования весовой функции при вычислении спектра. Лишь мгновенные оценки формантных частот по методу нулей сигнала чётко выявили разрывы формант на переходных участках в звукосочетаниях.

В этом примере исходным материалом для метода нулей являлись отфильтрованные сигналы с фильтрами в характерных диапазонах формантных частот для гласных русского языка. В каждый момент времени параллельно выполнялись оценки по фильтрам, соответствующим формантам гласных /*И*/ и /*А*/. Выбирались оценки того набора фильтров, в котором сумма пиков огибающей по всем трём формантам была наибольшей. Без такого отбора разброс оценок формантных частот слишком велик, и никакое сглаживание не улучшает поведения формантных треков. В частности, если исходить из обычного предположения, что следующее значение частоты некоторой форманты должно находиться как можно ближе к предыдущему, то в области переходного процесса произойдёт перескок оценок второй форманты на третью форманту.

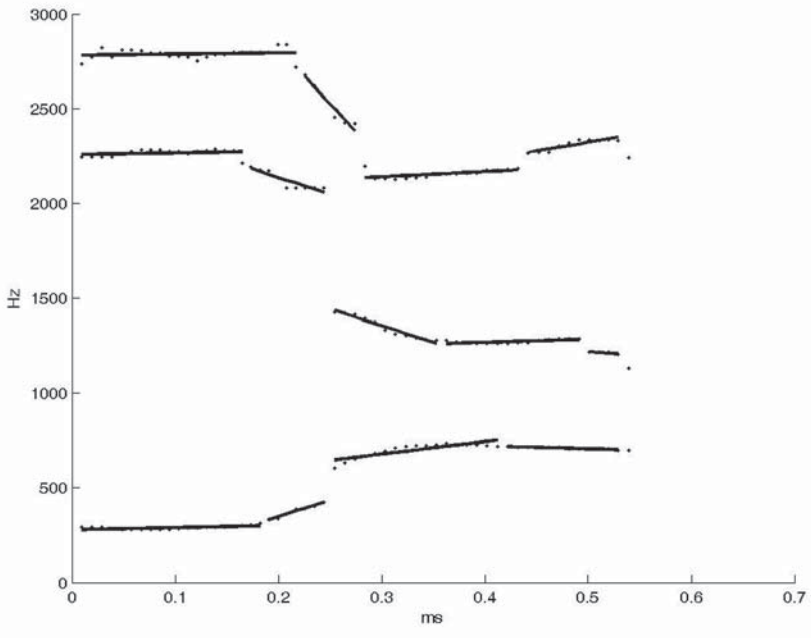


Рис. 3. Мгновенные оценки формантных частот в слоге /ИА/ (···) и кусочно-линейная аппроксимация треков (——)

Успех этого численного эксперимента позволяет сформулировать ещё один способ стабилизации оценок формантных частот, отличный от использования метода линейного предсказания в качестве предварительной оценки. Для каждого языка можно найти небольшое число характерных векторов формантных частот, примерно соответствующих гласным этого языка в том смысле, как их определяют фонетисты. Методика поиска этих характерных векторов путём кластеризации множества измерений формантных частот была описана в [16]. Распределение вероятностей каждого из этих характерных векторов может быть использовано для построения согласованных фильтров. Сигналы на выходе каждого набора фильтров подвергаются анализу согласно некоторому критерию, и оценки формантных частот выбираются для того набора фильтров, где значение этого критерия наилучшее. В частности, этот критерий может состоять в суммарной энергии — так, как это было применено в описанном выше примере.

Очевидно, что такой метод будет лучше всего работать на квазистационарных участках речевого сигнала, тогда как переходные процессы могут оцениваться с большей погрешностью. Однако можно сформировать алгоритм коррекции оценок на переходных процессах, используя устойчивые оценки формантных частот на краях переходного процесса.

Преимущество этого подхода заключается в том, что его можно применять для произвольного контекста, не заботясь о предварительной оценке положения во времени гласноподобных сегментов. При этом полностью используется информация о формантных образах гласных звуков в конкретном языке. Как было показано в данной работе и в предыдущих исследованиях на эту тему [5, 6], без учёта этой информации невозможно сколько-нибудь устойчивое определение формантных частот в речевом сигнале. Ещё одно преимущество заключается в подавлении колебаний, проникающих из подсвязочной области в речевой тракт. Это особенно важно при решении обратной задачи с целью определения формы речевого тракта, для чего нужно быть уверенным в том, что измеренные частоты действительно соответствуют резонансным частотам речевого тракта.

Заметим, что при таком подходе формантный анализ речевого сигнала становится зависимым от конкретного языка, его артикуляторного строя и формантных образов основных гласных. Интуитивно это представляется вполне оправданным. Это также объясняет неудачу многочисленных попыток построить универсальный устойчивый алгоритм определения формантных частот в речевом сигнале независимо от языка. Отсюда можно предположить, что и автоматический анализ взрывных согласных, назальных и фрикативных звуков также должен производиться с использованием специфических акустических свойств конкретного языка. Ясно, что основная трудность при этом заключается в создании достоверной базы акустических характеристик каждого языка на основе более или менее абстрактных методов анализа и ручной обработке полученных данных.

Заключение

Метод нулей сигнала характеризуется значительно меньшим разбросом оценок формантных частот в зависимости от типа регистрирующего микрофона и устойчив к шумам, особенно в низкочастотной области. Мгновенные оценки формантных частот этим методом на периоде основного тона могут быть уточнены (скорректированы) путём использования информации о типе гласного. Предположение о непрерывности формантных

треков при коррекции не оправдано. Поэтому коррекцию оценок формант следует выполнять путём кусочно-линейной аппроксимации с возможными разрывами треков.

ПРИЛОЖЕНИЕ

Экспериментально было установлено, что на оценку формантных частот влияет частота основного тона, причём это влияние особенно заметно сказывается при определении низкочастотных формант. Механизм этого явления был не вполне ясен. Ниже приводятся две простые математические модели, позволяющие изучить воздействие источника возбуждения на спектр сигнала и качественно описать соответствующие эффекты, которые проявляются при формантном анализе.

Представим речевой тракт как совокупность не связанных осцилляторов с собственными частотами F, определяющими форманты. Будем сначала считать, что эти осцилляторы колеблются без затухания под действием гармонического источника возбуждения с частотой основного тона F_0. Математически запишем это в виде задачи Коши для колебания $y(t)$:

$$y''+\omega^2 y = A\sin\Omega t,\ \ y(0)=y'(0)=0\,.$$

Здесь $\omega = 2\pi F$ — собственная (круговая) частота осциллятора, а $\Omega = 2\pi F_0$. Нетрудно видеть, что

$$y(t) = -\frac{A}{\omega^2-\Omega^2}\left(\frac{\Omega}{\omega}\sin\omega t - \sin\Omega t\right).$$

Преобразуем найденное решение, вводя величину $\varphi(t)\in[0,2\pi)$ — решение системы уравнений

$$\cos\varphi(t) = \frac{\Omega}{\omega}\sin(\omega+\Omega)t,\ \ \sin\varphi(t) = 1+\frac{\Omega}{\omega}\cos(\omega+\Omega)t\,,$$

а также числа

$$D = -\frac{A}{\omega^2-\Omega^2}\sqrt{1+\left(\frac{\Omega}{\omega}\right)^2}\,,\ \ \mu = 2\frac{\Omega}{\omega}\left[1+\left(\frac{\Omega}{\omega}\right)^2\right]^{-1}.$$

В итоге оказывается, что

$$y(t) = D\left[1+\mu\cos(\omega+\Omega)t\right]^{1/2}\cos\left[\Omega t+\varphi(t)\right].$$

Это решение легко интерпретируется при $\mu \ll 1$, то есть для формантных частот, много больших частоты основного тона. В этом случае

$$\varphi(t)\approx\frac{\pi}{2},\ D\approx-\frac{A}{\omega^2}\ \text{и}$$

$$y(t)\approx\frac{A}{\omega^2}\sin\Omega t+\frac{A\mu}{2\omega^2}\cos(\omega+\Omega)t\cdot\sin\Omega t\,, \qquad (1)$$

так что движения осцилляторов представляют собой колебания основного тона, на которые наложены колебания со сдвинутой формантной частотой, промодулированные колебаниями с частотой основного тона. Таким обра-

зом, рассмотренная простейшая модель качественно предсказывает не только модуляцию амплитуд формант из-за воздействия голосового источника возбуждения, но и их сдвиг в сторону увеличения частоты.

Влияние гармоник основного тона на спектр колебаний в речевом тракте можно качественно изучить и для источника более общего вида с учётом затухания собственных колебаний. Полагая, что голосовой источник $f(t)$ — это кусочно-гладкая периодическая функция, разложим её в ряд Фурье на периоде колебаний (например, в ряд Фурье по синусам, если $f(0) = 0$). Тогда задачу определения вынужденных колебаний осциллятора можно записать в виде

$$y''+2\gamma\omega y'+\omega^2 y = f(t) = \sum_{n=1}^{\infty} b_n \sin n\Omega t \quad (0<\gamma<1), \qquad (2)$$
$$y(0) = y'(0) = 0.$$

Её решение представимо как $y(t) = \sum_{n=1}^{\infty} y_n(t)$, где слагаемые находятся из задачи Коши:

$$y_n''+2\gamma\omega y_n'+\omega^2 y_n = b_n \sin n\Omega t, \;\; y_n(0) = 0, \; y_n'(0) = 0.$$

Можно вычислить, что $y_n(t) = y_n^{(0)}(t) + y_n^{(1)}(t)$, где

$$y_n^{(0)}(t) = b_n \left[\frac{e^{-\gamma\omega t} \sin\left(\sqrt{1-\gamma^2}\,\omega t\right) n\Omega\left(\Omega^2 + 2\gamma^2\omega^2 - \omega^2\right)}{\sqrt{1-\gamma^2}\,\omega\left(4\omega^2\Omega^2\gamma^2 + (\Omega^2-\omega^2)^2\right)} + \frac{2e^{-\gamma\omega t}\cos\left(\sqrt{1-\gamma^2}\,\omega t\right)\gamma\omega\Omega}{\left(4\omega^2\Omega^2\gamma^2 + (\Omega^2-\omega^2)^2\right)} \right]$$

$$y_n^{(1)}(t) = b_n \frac{(\omega^2 - n^2\Omega^2)\sin n\Omega t - 2\gamma n\omega\Omega \cos n\Omega t}{(\omega^2 - n^2\Omega^2)^2 + 4\gamma^2\omega^2 n^2\Omega^2}.$$

Решение задачи (2) $y(t) = \sum_{n=1}^{\infty} y_n(t) = \sum_{n=1}^{\infty} y_n^{(0)}(t) + \sum_{n=1}^{\infty} y_n^{(1)}(t) \equiv y^{(0)}(t) + y^{(1)}(t)$

интерпретируется так: в голосовом тракте существуют не только затухающие собственные колебания $y^{(0)}(t)$, но и колебания $y^{(1)}(t)$, которые определяются частотой основного тона. Это верно даже на интервале закрытой голосовой щели, т.е. для временных интервалов, где $f(t) = 0$. Поэтому при формантном анализе на интервале закрытой голосовой щели на получаемый результат влияет член сигнала

$$y^{(1)}(t) = \sum_{n=1}^{\infty} b_n \frac{(\omega^2 - n^2\Omega^2)\sin n\Omega t - 2\gamma n\omega\Omega \cos n\Omega t}{(\omega^2 - n^2\Omega^2)^2 + 4\gamma^2\omega^2 n^2\Omega^2} =$$

$$= \frac{1}{\omega^2}\sum_{n=1}^{\infty} b_n \frac{(1 - n^2\beta^2)\sin n\Omega t - 2\gamma n\beta \cos n\Omega t}{(1 - n^2\beta^2)^2 + 4\gamma^2 n^2\beta^2} =$$

$$= \frac{1}{\omega^2}\sum_{n=1}^{\infty} b_n \sin(n\Omega t + \varphi_n), \qquad (3)$$

где φ_n есть главное решение системы уравнений

$$\cos\varphi_n = \frac{(1-n^2\beta^2)}{(1-n^2\beta^2)^2+4\gamma^2n^2\beta^2},\ \sin\varphi_n = -\frac{2\gamma n\beta}{(1-n^2\beta^2)^2+4\gamma^2n^2\beta^2},\quad \beta=\frac{\Omega}{\omega}.$$

Слагаемое (3) искажает спектр собственных частот сигнала, в котором в итоге появляются колебания с частотами $n\Omega$. При небольших n частоты $n\Omega$ могут быть сравнимы с формантными. Амплитуды Фурье-гармоник функции (3) суть изменённые в $1/\omega^2$ раз амплитуды гармоник источника $f(t)$. Поэтому искажения оценок формантного анализа наиболее существенны для низких частот, когда отношение $1/\omega^2$ велико. Ещё раз подчеркнём, что сделанные выводы справедливы для любой кусочно-гладкой формы источника возбуждения.

Проведённый анализ объясняет причины экспериментально установленной зависимости оценок формантных частот от частоты основного тона источника голосового возбуждения, которая наблюдается для любых методов формантного анализа.

Литература

1. G.K. Vallabha, B.Tuller (2002). Systematic errors in formant analysis of steady-state vowels. Speech Communication, v.38, pp.141–160.

2. Цемель Г.И. Опознавание речевых сигналов. М.: Наука, 1971.

3. R.J. Niederjohn, M.Lahat (1985). A zero-crossing consistency method for formant tracking of voiced speech in high noise levels. IEEE on Acoustics, Speech and Signal Processing, ASSP–33, N2, 349–355.

4. Th.Sreenivas, R.J. Niederjohn (1992). Zero-crossing based spectral analysis and SVD spectral analysis for formant frequency estimation in noise, IEEE transactions on Signal Processing, v.40, N2, 282–293.

5. Сорокин В.Н., Трифоненков И.П. Об автокорреляционном анализе речевых сигналов. 1996. Акуст. ж., Т. 42. №3. С. 368–374.

6. Леонов А.С., Сорокин В.Н. К анализу резонансных частот речевого тракта. Информационные процессы, Т. 7. 2007. №4, 386–400. www.jip.ru.

7. K.Mystafa, I.C. Bruce (2006). Robust formant tracking for continuous speech with speaker variability. IEEE transactions on Audio, Speech, and Language Processing, v.14, N2, 435–444.

8. Сорокин В.Н., Макаров И.С. Распознавание пола диктора по голосу. Акустический ж. 2008. Т. 54, №4, С. 1–9.

9. A.El-Jaroudi, J.Makhoul (1991). Discrete All-Pole Modeling. IEEE Trans. Signal Process., vol.39, No.2, pp.411–423.

10. Сорокин В.Н., Макаров И.С. Обратная задача для голосового источника. Информационные процессы. 2006. Т. 6, №4, 375–395. www.jip.ru.

11. S.McCandless. An Algorithm for Automatic Formant Extraction Using Linear Prediction Spectra. // IEEE Trans. Acoust., Speech, Signal Process., vol.ASSP–22, 1974, pp.135–141.

12. G.Kopec. Formant Tracking Using Hidden Markov Models and Vector Quantization. // IEEE Trans. Acoust., Speech, Signal Process., vol.ASSP–34, 1986, pp.709–729.

13. D.Talkin. Speech Formant Trajectory Estimation Using Dynamic Programming with Modulated Transition Costs. // J.Acoust. Soc. Amer. S1, 1987, p.S55.

14. K.Xia, C.Espy-Wilson. A New Strategy of Formant Tracking Based on Dynamic Programming. // Proc. Int. Conf. Spoken Lang. Process., 2000, pp.55–58.

15. X.Chi, M.Sonderegger (2007). Subglottal coupling and its influence on vowel formants. Journal. Acoust. Soc. Am., v.122, N3, 1735–1745.

16. Сорокин В.Н., Цыплихин А.И. Сегментация и распознавание гласных. Информационные процессы, 2004. Т. 4. №2, С. 202–220.www.jip.ru.

В.Н. Сорокин,
доктор физико-математических наук,
ведущий научный сотрудник
Института проблем передачи информации РАН.
E-mail: vns@iitp.ru.

А.С. Леонов,
доктор физико-математических наук,
профессор кафедры математики
Московского инженерно-физического института
(Федеральный исследовательский ядерный университет).
Специалист в области решения обратных и некорректно
поставленных задач науки и техники
(обратные задачи теплопроводности и диффузии,
задачи обработки изображений,
задачи оптимального синтеза технических систем,
обратные задачи речевых технологий и др.).
Автор монографий по решению нелинейных некорректных задач.

И.С. Макаров,
Институт проблем передачи информации,
Российская академия наук, Москва, Россия.

Темпоральные характеристики интонации речи дикторов телевидения

С.Б. Жемерова
Санкт-Петербургский государственный университет

В данной работе рассматриваются темпоральные характеристики интонации: темп речи, паузы, а также длительность интонационных единиц в речи дикторов в новостных телепередачах. Новизна исследования определяется тем, что число описаний интонации в речи дикторов телевидения сравнительно невелико. Кроме того, приводимые исследователями данные достаточно противоречивы. Полученные результаты можно использовать для усовершенствования существующих систем синтеза и распознавания речи.

Введение

Речь дикторов телевидения — интересный предмет исследования благодаря её особой роли в языковом сообществе. С одной стороны, носители языка ожидают, что она должна быть нормативной, потому что её источник — средство массовой информации. С другой стороны, дикторы оказывают на норму значительное влияние именно по той причине, что их речь воспринимается как безоговорочно нормативная и её регулярно слышит большая часть языкового сообщества. Механизмы влияния радио и телевидения на формирование и распространение языковой нормы описаны многими исследователями [1:52, 2:44].

В качестве предмета настоящего исследования была выбрана именно речь профессиональных дикторов телевидения в новостных телепередачах.

В статье рассматриваются темпоральные характеристики интонации дикторов телевидения: темп речи, паузы, а также длительность интонационных единиц. Целью работы являлось создание индивидуальных речевых портретов дикторов телевидения, их сопоставление между собой и выявление существующих закономерностей.

Материалом для исследования послужили записи выпусков новостей. Для исследования были выбраны два крупнейших российских телеканала: «Первый канал» и канал «Россия».

На каждом из каналов было выбрано по два диктора — один мужчина и одна женщина, в исполнении которых на момент сбора материала было доступно большее количество записей:
— диктор Д. — «Первый канал», мужчина, 23 года, дикторского образования не имеет;
— диктор К. — «Первый канал», женщина, 47 лет, имеет профессиональное дикторское образование (курсы для работников радио и телевидения);
— диктор Л. — канал «Россия», женщина, 30 лет, дикторского образования не имеет;
— диктор М. — канал «Россия», мужчина, 36 лет, имеет высшее журналистское образование.

Для исследования было отобрано по две минуты записей каждого диктора. Общее время звучания дикторской речи составило 7 минут 367 миллисекунд.

Для каждого диктора подсчитывались:
средний темп речи;
средняя длина синтагмы в слогах и в миллисекундах;
процент синтагм, содержащих паузы, средняя длина внутрисинтагменной паузы;
процент границ синтагм, оформленных паузами, средняя длина межсинтагменной паузы;
количество пауз на минуту речи;
длительность пауз на секунду речи (отдельно отмечались случаи, когда последнее слово или словосочетание синтагмы было отделено паузой);
количество слогов в первой и во второй половинах синтагмы.

Для определения статистической значимости разницы в количестве слогов использовался коэффициент корреляции Пирсона. Для проверки гипотезы о зависимости темпа речи от длины синтагмы также использовался коэффициент корреляции Пирсона.

Темп речи

Таблица 1

Средний темп речи у разных дикторов (по всему материалу)[1]

	Темп (слогов в секунду)			
	минимум	**максимум**	**среднее**	**разброс**
Диктор Д.	7,02	7,49	7,23	0,47
Диктор К.	6,87	7,56	7,18	0,69
Диктор Л.	6,79	7,45	7	0,66
Диктор М.	6,24	7,51	7,84	1,27

[1] Здесь и далее в таблицах в графе «минимум» представлено минимальное среднее значение из встретившихся, в графе «максимум» — максимальное среднее значение, в графе «среднее» — общее среднее значение по всем текстам, в графе «разброс» — величина разброса значений (приводится в виде разницы между максимальным и минимальным значениями, так как данных для подсчёта стандартного отклонения недостаточно).

В таблице 1 представлены средние значения темпа речи для разных текстов. Как видно из таблицы, темп речи у разных дикторов различается незначительно. У дикторов-женщин средний темп несколько ниже, чем у мужчин. Максимальный темп речи несколько выше у дикторов К. и М. Это может быть связано с рядом причин: эти дикторы старше, чем дикторы Д. и Л., и у них несколько больший опыт работы в данной сфере. Кроме того, как уже было упомянуто, у диктора К. есть профессиональное дикторское образование, а у диктора М. — высшее журналистское, в ходе которого он мог получать дикторскую подготовку (к сожалению, доподлинно выяснить, так это или нет, не удалось). С каким именно из этих факторов связано различие в темпе речи, на имеющемся материале определить невозможно.

Максимальный разброс в значениях среднего темпа речи наблюдается у диктора М. Это связано, по всей видимости, с тем, что у данного диктора записано больше текстов, чем у других дикторов, и эти тексты достаточно разнообразны по тематике. Можно также предположить, что вариативность темпа речи, как и максимальное его значение, связаны с уровнем профессиональной подготовки диктора, однако данных, для того чтобы говорить об этом с уверенностью, недостаточно.

По данным О.Ф. Кривновой, средняя длина слога для среднего темпа речи в русском языка составляет 150–210 мс, что соответствует темпу речи 4,76–6,67 слогов в секунду [3: 40]. Таким образом, имеющиеся данные позволяют охарактеризовать темп речи дикторов в большинстве записей, скорее, как высокий по сравнению со средним для языка. Это соответствует данным разных исследователей [4; 5: 53], которые характеризуют темп речи дикторов телевидения как ускоренный по сравнению с нейтральной речью.

Надо заметить, что во многих других языках дикторам телевидения несвойственен быстрый темп речи. Так, в финском языке средний темп речи дикторов телевидения составляет 6,5 слога в секунду, в немецком — 5,9, в английском — 5,4 слога в секунду [6: 383]. Этот темп речи не является ускоренным по сравнению с нормативным. К примеру, для английского языка В.Левельт приводит среднее значение темпа речи 5–6 слогов в секунду [7: 306], а Дж. Лэйвер — 5–5,5 слогов в секунду [8: 541].

Если говорить о связи темпа речи с функциями языка, то Т.М. Надеина отмечает: быстрый темп речи приводит к тому, что информация воспринимается как менее понятная, актуальная и интересная, но содержание оценивается как более динамичное [9: 157]. Таким образом, можно сделать вывод, что, говоря в темпе, быстром по сравнению со среднеязыковым, дикторы стремятся сделать свою речь более динамичной и тем самым усилить функцию воздействия.

Длина синтагмы

Таблица 2

Средняя длина синтагмы в слогах у разных дикторов (по всему материалу)

	Слогов в синтагме			
	минимум	**максимум**	**среднее**	**разброс**
Диктор Д.	6,56	7,87	8,5	1,94
Диктор К.	7,65	9	10,56	2,91
Диктор Л.	7,13	8,53	10	2,87
Диктор М.	7,25	9,19	10,22	2,97

Таблица 3

Средняя длина синтагмы в миллисекундах у разных дикторов (по всему материалу)

	Длина синтагмы (мс)			
	минимум	**максимум**	**среднее**	**разброс**
Диктор Д.	933	1093	1181	248
Диктор К.	1112	1273	1540	428
Диктор Л.	1063	1219	1353	290
Диктор М.	1070	1297	1575	505

Средняя длина синтагмы как в слогах, так и в миллисекундах, наибольшая у дикторов К. и М. Вызвано это, вероятно, теми же причинами, что и наибольший максимальный темп в их речи.

У этих же дикторов наблюдается наибольший разброс значений длины синтагм. О.А. Прохватилова отмечает, что синтагматическому членению в информационных текстах принадлежит особая роль, «поскольку именно делимитация речевого потока позволяет максимально актуализировать содержательные компоненты высказываний, обозначить смысловые центры» [5: 51]. Таким образом, возможно, что большой диапазон длины синтагм в речи данных дикторов свидетельствует о большем умении или стремлении пользоваться синтагматическим членением как выразительным средством. Это вполне соотносится с тем фактом, что у этих двух дикторов имеется специальное журналистское или дикторское образование, а также наибольший опыт работы в данной сфере.

О.Ф. Кривнова отмечает, что при среднем темпе произнесения в русском языке длина синтагм составляет около 1100–1500 мс [3: 29]. По данным Н.Б. Вольской, длина синтагмы в спонтанной речи колеблется от 1 до 1,7 секунды [10: 133], а в чтении — от 0,8 до 1,2 секунды [11: 169]. Схожие результаты приводят Л.В. Бондарко, Н.Б. Вольская, С.О. Тананайко и Л.А. Васильева. По их данным, средняя длина синтагмы в спонтанной

речи составляет 1,14 секунды, а в чтении — 1,12 секунды [12]. Таким образом, можно говорить о том, что для русского языка средняя длина синтагм в речи дикторов не отличается от нормативной.

Что касается длины синтагмы в слогах, то Н.Б. Вольская приводит данные, согласно которым средняя длина синтагмы при чтении в русском языке составляет 7,2–8,1 слога [13]. Данные значения несколько ниже, чем полученные в настоящей работе. Связано это, очевидно, с темпом речи.

У всех дикторов наблюдается статистически значимая корреляция между длиной синтагмы в слогах и темпом речи: чем длиннее синтагма, тем выше темп. Что касается темпа речи внутри синтагмы, то у всех дикторов наблюдается статистически значимое замедление темпа речи от начала синтагмы к концу. Как уже было упомянуто выше, это является следствием т. н. предпаузального удлинения и не является чертой, характерной только для дикторов телевидения.

Паузы

Таблица 4

Средний процент границ синтагм, оформленных паузами, у разных дикторов (по всему материалу)

	Границ синтагм оформлено паузами			
	минимум	**максимум**	**среднее**	**разброс**
Диктор Д.	17,65%	27,27%	22,22%	9,62%
Диктор К.	21,05%	52,63%	48,10%	31,58%
Диктор Л.	45,83%	72,73%	53,93%	26,90%
Диктор М.	9,09%	23,81%	16,90%	14,72%

Как видно из таблицы, у дикторов-женщин паузы между синтагмами встречаются значительно чаще, чем у мужчин.

Надо заметить, что значения, встретившиеся у дикторов-женщин, близки к нормативным для русского языка. Так, Л.В. Бондарко, Н.Б. Вольская, С.О. Тананайко и Л.А. Васильева приводят данные, согласно которым в русском языке в спонтанной речи паузами оформлено в среднем 53,8% синтагм, а в чтении — 65,6% [12]. По данным Н.Б. Вольской, в русском языке в спонтанной речи паузами оформлены 57% синтагм, а в чтении — 53% [10: 169].

Отметим, что, по данным Т.М. Надеиной, при большем количестве пауз между синтагмами содержание текста оценивается как более яркое, разнообразное, интересное, украшенное, но менее полезное и понятное [9: 157]. Таким образом, можно предположить, что дикторы-женщины стремятся сделать свою речь более выразительной, а мужчины — более информативной.

Таблица 5

Средняя длина паузы между синтагмами у разных дикторов (по всему материалу)

	Средняя длина межсинтагменной паузы (мс)			
	минимум	**максимум**	**среднее**	**разброс**
Диктор Д.	266	421	348	154
Диктор К.	251	356	289	105
Диктор Л.	241	301	267	60
Диктор М.	297	398	337	101

Как видно из таблицы, средняя длина пауз между синтагмами больше у дикторов-мужчин.

Для сравнения: по данным Л.В. Бондарко, Н.Б. Вольской, С.О. Тананайко и Л.А. Васильевой, средняя длина пауз в русском языке в спонтанной речи составляет 496 мс, а в чтении — 514 мс [12]. Таким образом, можно говорить о том, что для речи дикторов телевидения — как спонтанной, так и при чтении — свойственны более короткие паузы, чем для обычной речи. Несмотря на то, что речь дикторов телевидения не является спонтанной, значения длительности пауз в ней ближе к значениям, свойственным спонтанной речи, чем к значениям, характерным для чтения.

Таблица 6

Процент синтагм, содержащих паузы, у разных дикторов (по всему материалу)

	Синтагм, содержащих паузы			
	минимум	**максимум**	**среднее**	**разброс**
Диктор Д.	0%	5,56%	1,05%	5,56%
Диктор К.	0%	33,33%	11,90%	33,33%
Диктор Л.	0%	12,00%	6,67%	12,00%
Диктор М.	0%	8,33%	1,30%	8,33%

Внутрисинтагменные паузы являются важным выразительным средством в художественной и публицистической речи. Так, Н.В. Черемисина-Ениколопова отмечает, что «такая психологическая, или выразительная, аффективная пауза предшествует важному слову и как бы готовит читателя (и слушателя) к восприятию этого слова: возникает напряжение, увеличивающее смысловой вес постпаузного слова» [14: 161]. Таким образом, можно сделать вывод о том, что, используя в своей речи внутрисинтагменные паузы, дикторы стремятся усилить воздействующую функцию языка.

Как видно из таблицы, паузы внутри синтагм встречаются у женщин значительно чаще, чем у мужчин. Таким образом, здесь также прослеживается уже упомянутая тенденция к тому, что женщины активнее стремятся сделать свою речь выразительнее.

Таблица 7

Средняя длина паузы внутри синтагмы у разных дикторов (по всему материалу)

	Средняя длина внутрисинтагменной паузы (мс)			
	минимум	максимум	среднее	разброс
Диктор Д.	267	267	267	0
Диктор К.	146	205	168	59
Диктор Л.	138	192	175	54
Диктор М.	108	108	108	0

Делать какие-либо выводы о средней длине пауз внутри синтагмы невозможно ввиду недостаточного количества материала: у дикторов-мужчин встретилось по одной паузе внутри синтагмы на весь материал.

Таблица 8

Среднее количество пауз на минуту речи у разных дикторов (по всему материалу)

	Пауз на минуту речи			
	минимум	максимум	среднее	разброс
Диктор Д.	10,32	13,88	12,17	3,56
Диктор К.	16,74	36,12	27,37	19,38
Диктор Л.	25,89	33,65	29,85	7,76
Диктор М.	4,43	10,27	7,82	5,84

Как видно из таблицы, женщины делают паузы в речи чаще, чем мужчины. Это очевидно из того факта, что женщины делают больше пауз как внутри синтагм, так и между ними.

Как известно, при чтении текста дыхательный ритм является одним из самых существенных физиологических факторов, которые потенциально могут оказывать влияние на паузацию. О.В. Кривнова приводит данные, согласно которым средняя частота дыхательных пауз в речи составляет 16–20 в минуту [15]. Таким образом, можно говорить о том, что количество пауз в речи дикторов-женщин находится в рамках нормы, в то время как у дикторов-мужчин количество пауз в речи ниже нормативного.

Надо заметить, что крайне низкие значения для диктора М. могут быть обусловлены тем, что в его исполнении были доступны самые короткие записи, длина которых не позволяет собрать статистику, достаточную для подсчёта среднего.

Выводы

Итак, проведённое исследование позволяет сделать определённые выводы о темпоральных характеристиках интонации в речи дикторов телевидения.

Средний темп речи дикторов телевидения колеблется от 7 до 7,8 слога в секунду. Такой темп можно охарактеризовать как быстрый. У разных дикторов темп речи различается незначительно, но можно отметить, что средний темп речи у дикторов-мужчин несколько выше, чем у дикторов-женщин. Также можно предположить, что максимальный темп речи выше у дикторов с более высоким уровнем профессиональной подготовки.

Средняя длина синтагмы в дикторской речи составляет от 1093 до 1297 мс. Данные значения практически совпадают со значениями этого параметра, приводимыми для русского языка разными исследователями. Таким образом, можно говорить о том, что средняя длина синтагмы в речи дикторов телевидения не отличается от среднеязыковой.

Средняя длина синтагмы в слогах в материале настоящего исследования несколько выше среднеязыковых значений. Это, очевидно, связано с темпом речи.

Наибольшая средняя длина синтагмы — как в слогах, так и в миллисекундах — наблюдается у дикторов с более высоким уровнем профессиональной подготовки.

У всех дикторов прослеживается статистически значимая корреляция между длиной синтагмы в слогах и темпом речи: чем длиннее синтагма, тем выше темп. Кроме того, у всех дикторов наблюдается статистически значимое замедление темпа речи от начала синтагмы к концу, однако это является следствием т. н. предпаузального удлинения и не является чертой, характерной только для дикторов телевидения.

В речи разных дикторов паузами оформлены в среднем от 17 до 54% синтагм. В речи дикторов-женщин паузы между синтагмами встречаются значительно чаще, чем в речи дикторов-мужчин. Надо заметить, что значения данного параметра, полученные в настоящем исследовании для дикторов-женщин, близки к среднеязыковым, в то время как в речи дикторов-мужчин наблюдаются значения более чем в два раза ниже среднеязыковых.

Средняя длина пауз между синтагмами в речи разных дикторов составила от 267 до 348 мс. Данные значения ниже значений, приводимых другими исследователями как для спонтанной речи, так и для чтения.

Паузы внутри синтагм у женщин встречаются значительно чаще, чем у мужчин. Средняя длина внутрисинтагменной паузы у разных дикторов составляет от 108 до 267 мс.

Количество пауз в речи разных дикторов колеблется от 9 до 30 в минуту. Средняя частота встречаемости пауз в речи дикторов-женщин находится в рамках нормы, обусловленной естественным дыхательным ритмом, в то время как у дикторов мужчин количество пауз в речи значительно ниже нормативного.

Хотя объём материала, использованного в данной работе, безусловно, не позволяет делать окончательных выводов о наличии тех или иных просодических особенностей в речи дикторов телевидения, однако и на имеющемся материале стабильно прослеживается множество закономерностей.

В целом, темпоральные характеристики интонации в речи дикторов телевидения незначительно отличаются от таковых в обычной устной разговорной речи. Так, полученные значения средней длины синтагмы и количества пауз в речи дикторов телевидения практически совпадают со значениями данных параметров, приводимыми разными исследователями для спонтанной речи и чтения.

К интонационным особенностям речи дикторов телевидения относятся, в первую очередь, темп речи, более высокий по сравнению с нормативным, а также меньшая средняя длина пауз между синтагмами.

Темпоральные особенности интонации в речи дикторов телевидения зависят от ряда факторов. В первую очередь, это пол диктора. У дикторов-женщин средний темп речи ниже по сравнению с дикторами-мужчинами. В речи дикторов-мужчин меньше пауз как внутри синтагм, так и между ними, а средняя длительность межсинтагменных пауз выше, чем у дикторов-женщин. Кроме того, полученные данные о количестве пауз внутри синтагм и между ними позволяют предположить, что для женщин более важным является сделать свою речь более выразительной, в то время как мужчины стремятся к большей информативности речи.

На темпоральные особенности интонации оказывает влияние уровень профессиональной подготовки диктора. Так, для дикторов с профессиональным дикторским образованием и большим опытом работы характерна большая вариативность темпа речи и длины синтагм, а также их максимальные значения.

Кроме того, можно предполагать, что на интонационные характеристики в речи диктора оказывает влияние тематика читаемого текста. Так, для текстов о культуре и спорте характерен более низкий темп речи и большая длина пауз между синтагмами. Кроме того, в них чаще встречается явление отделения паузой последнего слова в тексте.

Необходимо также отметить, что интонационные средства, используемые дикторами телевидения, позволяют предполагать, что первичной в их речи является воздействующая функция языка.

Литература

1. Frazer T.C. «Heartland» English: Variation and Transition in the American Midwest. — Tuscaloosa, London: The University of Alabama Press, 1993.

2. Беликов В.И., Крысин Л.П. Социолингвистика. М.: Рос. гос. гуманит. ун-т, 2001.

3. Кривнова О.Ф. Ритмизация и интонационное членение текста в «процессе речи-мысли» (опыт теоретико-экспериментального исследования), Автореф. дисс. д. филол. наук. М.: МГУ, 2007.

4. Гришина О.А. Просодические особенности речи красноярских дикторов. Красноярск, 2001.

5. Прохватилова О.А. Фоностилистика: стилистический анализ звучащей речи: учеб.-мет. пособие. Волгоград: Изд-во Волгоградского гос. ун-та, 1996.

6. Ilvonen A. [et al.] Comparison of Prosodic Characteristics in English, Finnish and German Radio and TV Newscasts. // Proceedings of The XIIIth International Congress of Phonetic Sciences, v.2. — Stockholm: Arne Strömbergs Grafiska, 1995. p.382–385.

7. Levelt W.J.M. Speaking: From Intention to Articulation. Cambridge: The MIT Press, 1995.

8. Laver J. Principles of Phonetics. Cambridge: Cambridge University Press, 1994.

9. Надеина Т.М. Функционирование просодических средств как факторов речевого воздействия // Фонетические чтения в честь 100-летия со дня рождения Л.Р. Зиндера. СПб.: Филологический факультет СПбГУ, 2004.–С.155–159.

10. Вольская Н.Б. О паузе и не только о ней // Фонетические чтения в честь 100-летия со дня рождения Л.Р. Зиндера. СПб.: Филологический факультет СПбГУ, 2004. С.129–136.

11. Вольская Н.Б. О паузах виртуальных и реальных // Проблемы и методы экспериментально-фонетических исследований. К 70-летию профессора кафедры фонетики и методики преподавания иностранных языков Л.В. Бондарко / Отв. ред. Н.Б. Вольская, Н.Д. Светозарова. СПб.: Филологический факультет СПбГУ, 2002. С.165–170.

12. Bondarko L.V. [et al.]. Phonetic Properties of Russian Spontaneous Speech // Proceedings of The XVth International Congress of Phonetic Sciences. Barcelona, 2003.

13. Volskaya N.B. Virtual and Real Pauses at Clause and Sentence Boundaries // Proceedings of The XVth International Congress of Phonetic Sciences. Barcelona, 2003.

14. Черемисина-Ениколопова Н.В. Законы и правила русской интонации: учеб. пособие. М.: Флинта: Наука, 1999.

15. Кривнова О.Ф. Паузирование при автоматическом синтезе речи / О.Ф. Кривнова, И.С. Чардин. М.: МГУ, 1999.

С.Б. Жемерова,
Санкт-Петербургский государственный университет

Компьютерные сетевые технологии в обучении лингвистическим дисциплинам (инновационные учебно-научные Интернет-порталы по русской фонетике)

Г.Е. Кедрова,
кандидат филологических наук

В.В. Потапов,
доктор филологических наук

А.М. Егоров

Е.Б. Омельянова

М.В. Волкова

Анализируется опыт создания Интенет-порталов «Русская фонетика» (URL: http://fonetica.philol.msu.ru/) и «Фонетика русских диалектов» (URL: http://dialect.philol.msu.ru), на основе которого рассматриваются базовые принципы конструирования мультимедийной интерактивной и адаптивной компьютерной обучающей среды по лингвистике.

Компьютеры и, особенно, глобальная компьютерная связь уверенно занимают доминирующие позиции среди мировых коммуникационных систем. Наиболее впечатляющие успехи достигнуты сегодня в области компьютерной поддержки обучения и образования — естественно, в первую очередь, в дистанционной их форме. Дистанционное обучение особенно актуально для России с её географической протяжённостью, специфической, уже достаточно давно сложившейся региональной системой образования. В немалой степени его актуальность обусловлена и новыми аспектами национальной образовательной доктрины, которая предполагает не только общедоступность качественного образования для населения страны, но и

создание условий для обучения и переобучения на протяжении всей активной жизни человека, т.н. *life-long learning*.

Считается, что существенную помощь в решении проблемы информационной поддержки образования и обучения могли бы оказать целенаправленно формируемые специалистами профессиональные научно-образовательные ресурсы и сервисы (так называемая сеть «Web-2.0 / Веб-2.0»). Такая сеть должна будет стать базой для эффективной подготовки специалистов вне школ, университетов и институтов, и именно она может служить полноценной основой для «продолжающегося», дополнительного, образования и обучения, программ повышения квалификации и переподготовки специалистов, столь востребованных во всех областях жизни современного общества.

В настоящий момент сфера Веб-2.0 активно разрабатывается и в России: в МГУ им. М.В. Ломоносова, других образовательных учреждениях, в институтах РАО и РАН, разнообразных коммерческих и некоммерческих образовательных учреждениях. Уже сейчас в этих организациях накоплен огромный информационный ресурс, специально подготовленный для образовательных целей, который включает электронные библиотеки (в том числе аудио- и видеолекции); активно формируются специализированные образовательные порталы, электронные справочные системы, онлайновые словари, учебно-справочные интегрированные гипермедийные комплексы; создаются электронные учебники, компьютерные тренажёры и симуляторы, а также системы администрирования и технологической поддержки учебного процесса в дистанционной форме [1].

Основные теоретические и методологические предпосылки формирования сети Web-2.0 послужили основой в конструировании элементов компьютерной обучающей среды, предназначенной для преподавания филологических дисциплин, на Веб-сайте Центра новых информационных технологий в гуманитарном образовании (ЦНИТ ГО) филологического факультета МГУ. Пилотные проекты, выполненные на сайте в русле инновационной концепции дистанционного обучения, базовым компонентом которого выступает распределённая компьютерная обучающая среда, — это Веб-порталы «Русская фонетика в Интернете» (URL: http://fonetica.philol.msu.ru/) и «Фонетика русских диалектов» (URL: http://dialect.philol.msu.ru).

Выбор этих учебных курсов продиктован изначально присущей этой области лингвистического знания гипермедийностью и междисциплинарным характером изучаемой информации. Хорошо известно, что эти курсы усваиваются студентами и учащимися с большим трудом, во многом, именно в силу разноплановости и многоформатности своего информационного наполнения. Поэтому структура и формат представления электронных учебных материалов в наших Интернет-порталах — объектно-ориентированные, т.е. предъявляемые пользователю Интернет-страницы формируются динамически при каждом запросе пользователя из сформированных a priori информационных элементов разной модальности и размерности, которые хранятся в базе данных и в дальнейшем, будучи определены в соответствии со стандартным метаязыком описания учебных информационных компонентов, могут быть неоднократно использованы в составе самых разных учебных курсов и информационно-справочных материалов энциклопедического характера [2].

Необходимо подчеркнуть, что этот подход предъявляет особые требования к технологиям конструирования учебного Интернет-пространства. В первую очередь, он заставляет максимально чётко и формализованно определять исходные принципы отбора и описания языкового материала, который будет положен в основу базового иллюстративного массива примеров и выстраивания на его основе структурированного описания всей

информационной области. Мы считаем, что успешное решение этой сложной задачи возможно, если в основу построения электронного учебника и сопутствующей системы электронных упражнений как базового компонента всякой обучающей среды положена индексированная и исчерпывающим образом откомментированная база языковых данных, иллюстрирующая все значимые противопоставления на каждом уровне языковой системы.

Рассмотрим подробнее принципы формирования такого типа базы данных, которая легла в основание Интернет-портала по русской фонетике (рис.1).

Использованная в основе обучающего гипертекстового пространства информационного портала по русской фонетике база данных была сформирована из единиц всех уровней русской звучащей речи (звук, слог, фонетическое слово, ритмическая группа, ритмомелодические единства). Все её элементы были проаннотированы не только в отношении заключённой в них информации, но и в соответствии с глобальными и контекстными задачами обучения (реализуемые через рекомендуемые схемы навигации по узлам надстраиваемого гипертекстового пространства) и задачей формирования полезных навыков (реализуемой через систему обучающих упражнений). Исходно все эти единицы были сгруппированы нами по принципу минимальных пар в кластеры. В информационном пространстве фонетического знания такие минимальные пары позволяют наглядно представить все функционально значимые в языке бинарные и многомерные оппозиции. При этом бинарные многомерные оппозиции поставляют основной материал для построения систем, поддерживающих процесс исследования гипертекстовой среды обучающих и контролирующих упражнений, а многомерные оппозиции вместе с пропорциональными позволяют выстроить основ-

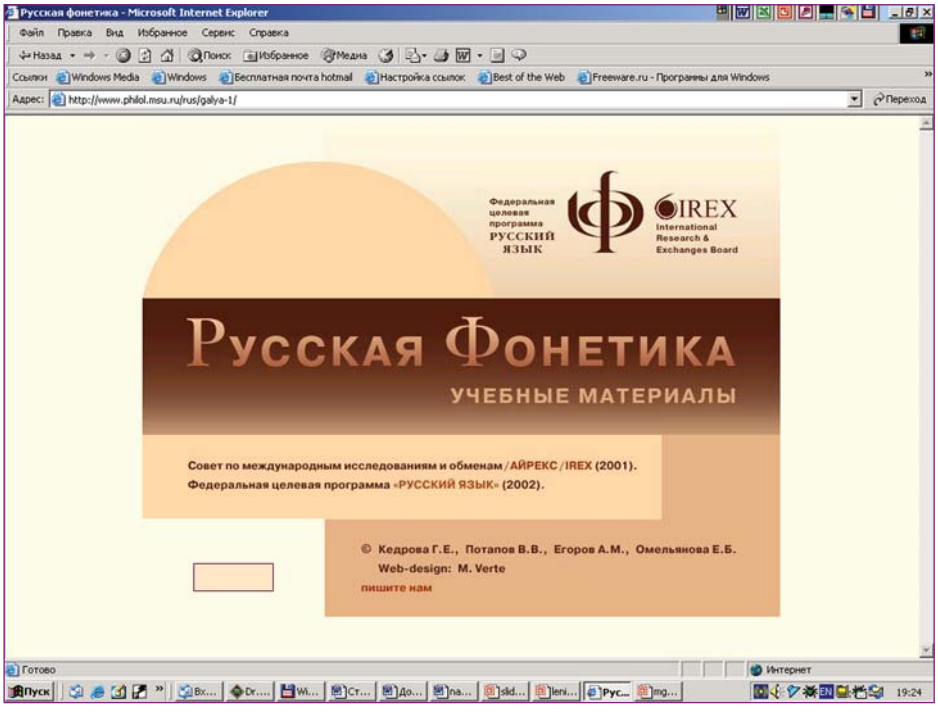


Рис. 1. Титульная страница Интернет-портала по русской фонетике

ные оси гипертекстового пространства, отражающие структурное взаиморасположение понятий, описывающих фонетическую систему языка. Необходимо также подчеркнуть, что благодаря введённому Н.С. Трубецким понятию нейтрализации структурное описание фонетического уровня языка естественным образом объединяется с представлениями об особенностях функционирования этой системы в речи, реальном речевом потоке.

Нам представляется, что только гипертекстовая технология формирования и представления знаний позволяет интегрировать эту составляющую в рамках единого многомерного когнитивного пространства. Крайне важно также и то, что такой подход изначально ориентирован на сохранение одного из ключевых параметров гипертекста как особого типа информационных структур, а именно: его открытости и множественности перспектив выстраивания иерархии понятий в рамках определённого знания [3]. Этот принцип определяет ещё одно требование к содержанию учебно-справочных Интернет-материалов (помимо собственно гипертекстовой оптимизации) — то, что все они построены на использовании только реально зафиксированного фактического языкового материала. Решение этих задач потребовало разработки специальных технологических решений и создания особых вспомогательных инструментальных средств. Так, для формирования экспериментальных фонетических баз данных, которые бы представляли все возможные фонетические минимальные смысловые пары в русском языке, была подготовлена специализированная компьютерная программа, которая позволяет из предварительно размеченного списка слов (с указанием места ударения) — компьютерного словаря любой размерности — генерировать базы данных (иллюстративных словарных единиц), обладающих заданными фонетическими параметрами [4]. Результатом работы программы стал компьютерный учебный фонетический словарь (69.000 единиц), или структурированная аннотированная база данных языковых фонетических примеров. Эта база является мультимедийной по определению и содержит для своих ключевых элементов как символьное представление (транскрипционный знак / текст / схема), так и соответствующие звуковые, анимационные и, при необходимости, видеофайлы. Такая база данных иллюстративных звуковых файлов на нашем Интернет-портале формировалась на основе аудиозаписей русской речи (мужской и женский голоса), отцифрованных и отсегментированных средствами условно свободно распространяемого программного пакета CoolEdit. База данных анимационных иллюстраций создавалась на основе эталонного банка кинофоторентгенограмм русской речи [5], видеофайлы (фронтальная видеосъёмка артикуляции русских звуков) редактировались покадрово в графическом пакете PageMaker с использованием программы формирования компьютерных анимаций Anigraph.

Подготовленная таким образом база данных по реализациям всех русских гласных и согласных звуков составляет более 5000 единиц. Длительность озвученных фрагментов (мужской и женский голоса) — 53 минуты. Весь звучащий массив был отцифрован и помещён для дальнейшего анализа на CD-ROM. Из массива озвученных слов были отобраны для акустического анализа 150 единиц. Этот материал включает минимальные пары и квазиминимальные пары слов по всем гласным и согласным фонемам русского языка во всех позициях в слове с учётом всех типов консонантного и вокального окружения.

Такой подход к конструированию электронных учебных материалов оказался достаточно трудоёмким. В итоге только по завершении базового этапа работ по созданию Веб-портала «Русская фонетика» были подготовлены 979 текстовых файлов, 53 анимационных файла (в формате avi), 98 графических файлов, 35 анимационных видеофайлов, 172 звуковых файла. В настоящее время все материалы, относящиеся к обучающей

среде по фонетике русского языка и выставленные в открытом доступе по указанному выше адресу на Интернет-сервере филологического факультета, занимают 53.291 KB, или 1.177 файлов.

Успешная реализация разработанных технологий позволила на следующем этапе провести апробацию этих методик при конструировании мультимедийного учебно-справочного и научного ресурса по курсу «Фонетика русских диалектов». Существенно, что этот Веб-ресурс адресован не только студентам для помощи в усвоении программного материала, но и преподавателям — для подготовки лекций и семинаров по предмету. Кроме того, в его задачи входило формирование представления о диалектах русского языка как элементе народной культуры, что может представлять особый интерес для самого широкого круга и российских, и иностранных пользователей. Такая апробация разработанных нами технологий прошла успешно, тем самым эти технологии и методики доказали, на наш взгляд, свою высокую эффективность.

В настоящее время Интернет-портал «Фонетика русских диалектов» включает 57 текстовых информационных модулей, 212 интерактивных мультимедийных упражнений, 122 контрольных теста, 2090 звуковых иллюстраций, 515 графических иллюстраций, 52 интерактивные диалектологические карты, 57 сонограмм звучащих примеров, 161 дефиницию терминов и терминологических понятий в Глоссарии.

В результате база текстовых и мультимедийных данных, которые сформировали информационное пространство учебника по русской фонетике, включает в себя следующие категории.

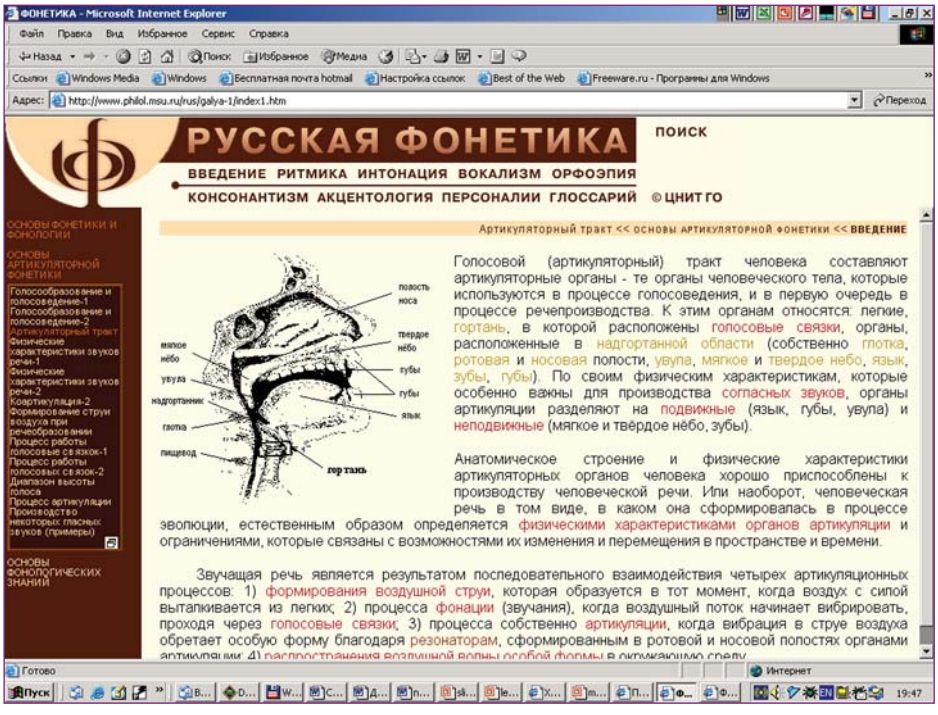


Рис. 2. Страница подраздела «Артикуляторный тракт»

1. Детальное и исчерпывающее описание в гипертекстовом режиме (текстовые файлы в HTML-формате) особенностей русской звуковой системы, интонологии и акцентологии, сформированное на основе аннотированного и индексированного словаря фонетических примеров; пример такого презентационного информационного блока показан на рис. 2.

2. Анимационное представление артикуляции русских звуков, выполненное на основе кинофоторентгенограмм реальной речи; элемент базы компьютерных анимаций показан на рис. 3.

3. Видеозаписи видимых артикуляционных движений (прежде всего работа губ); элемент базы компьютерных видеоанимаций показан на рис. 4.

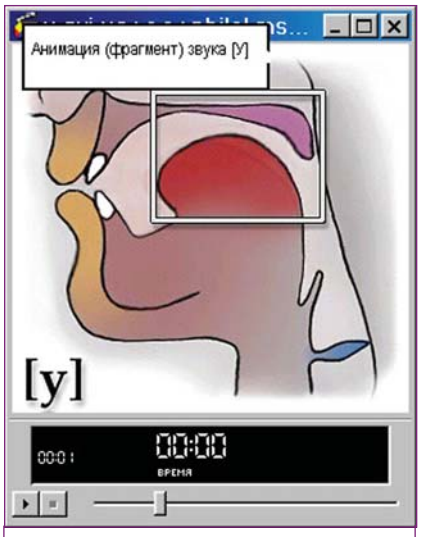


Рис. 3. Анимационная иллюстрация артикуляции гласного [у]

4. Схемы и графики, отражающие существенные параметры разных фонетических понятий и представлений; пример схематического представления особенностей ритмической структуры русского слова в сочетании с его звучанием показан на рис. 5a, пример схемы, иллюстрирующей подвижное ударение в русском глагольном словоизменении, показан на рис. 5b.

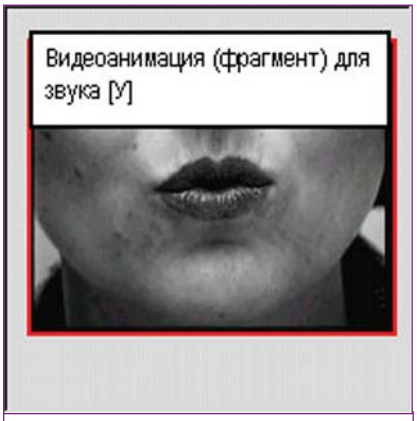


Рис. 4. Анимационная видеоиллюстрация губной артикуляции

5. Звуковые файлы, демонстрирующие реализацию звуков, слогов и фонетических слов в речи; пример отображения звуковых иллюстраций представлен на рис. 6 (звучащие иллюстрации отмечены в тексте зелёным цветом).

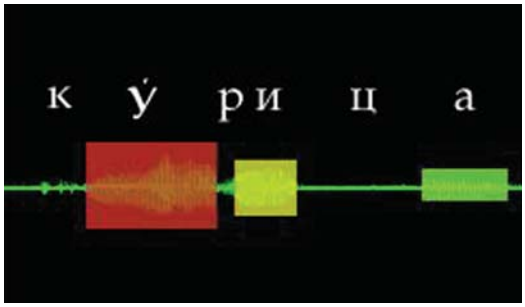
6. Контекстные выпадающие окна с дефинициями терминов и терминологических понятий, встречающихся в тексте; пример терминологического справочного окна представлен на рис. 7.

7. Акустические характеристики (осциллограмма, спектр, огибающая тона) речевых сегментов разной размерности; примеры отображения акустических характеристик звучащего иллюстративного материала представлены на рис. 8 и 9.

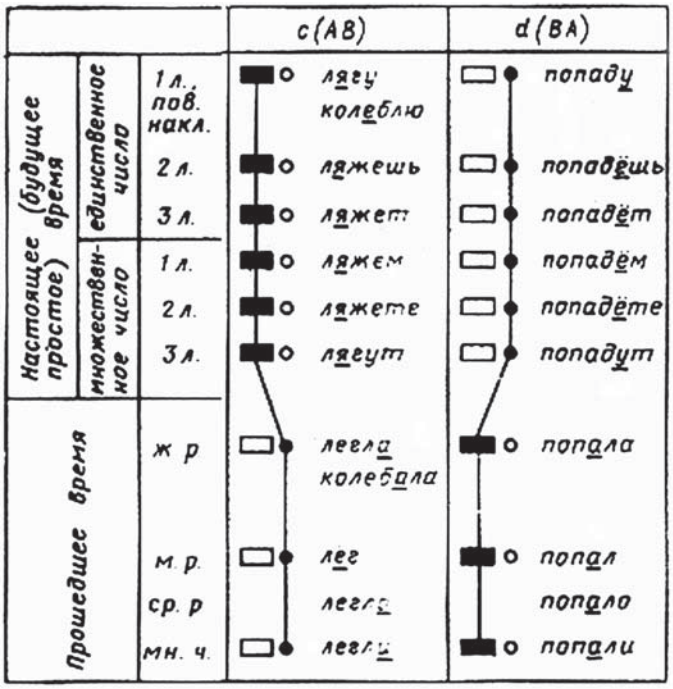
8. Графические иллюстрации (диалектологические карты); пример их презентации представлен на рис. 10.

В результате выполнения проекта «Русская фонетика» в Интернете размещены следующие функциональные модули гипертекстовой образовательной среды:

- вводные материалы по артикуляторной и акустической фонетике, методам структурного описания языка;
- русская произносительная база (артикуляторно-перцептивный аспект);



(a)



(b)

Рис. 5. Изображение ритмической структуры слова «курица» (a), ритмические схемы глагольного словоизменения (b)

- система русских гласных звуков;
- русский вокализм с теоретической точки зрения;
- русский консонантизм;
- русская ритмика;
- русская акцентуация;
- русская интонация;
- просодия русской речи;
- исторический очерк русской орфоэпии;
- современная орфоэпия;

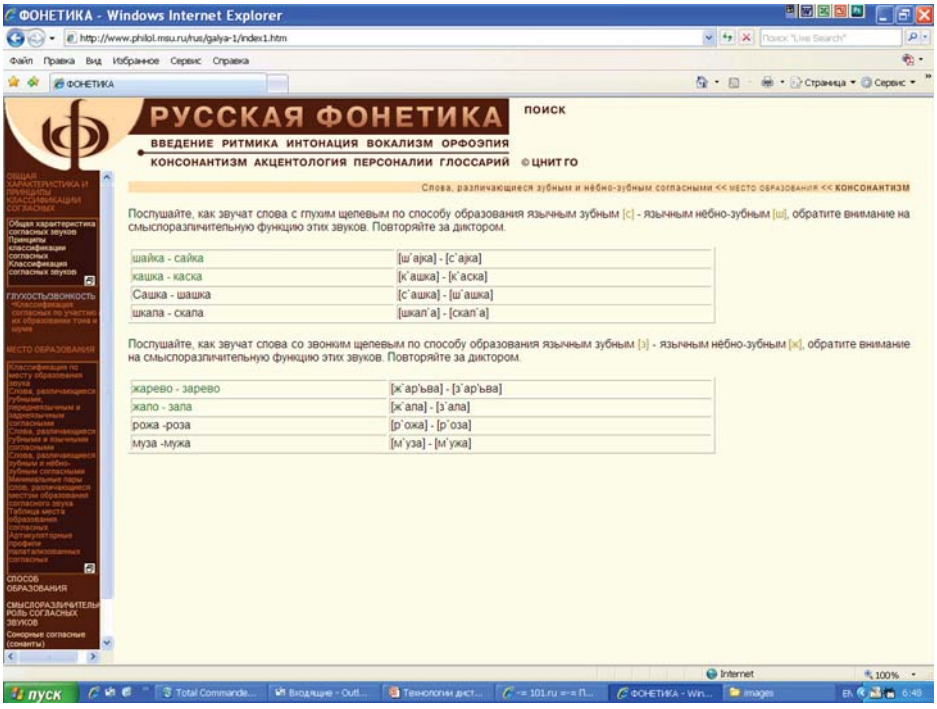


Рис. 6. Страница подраздела звучащих примеров — минимальных пар

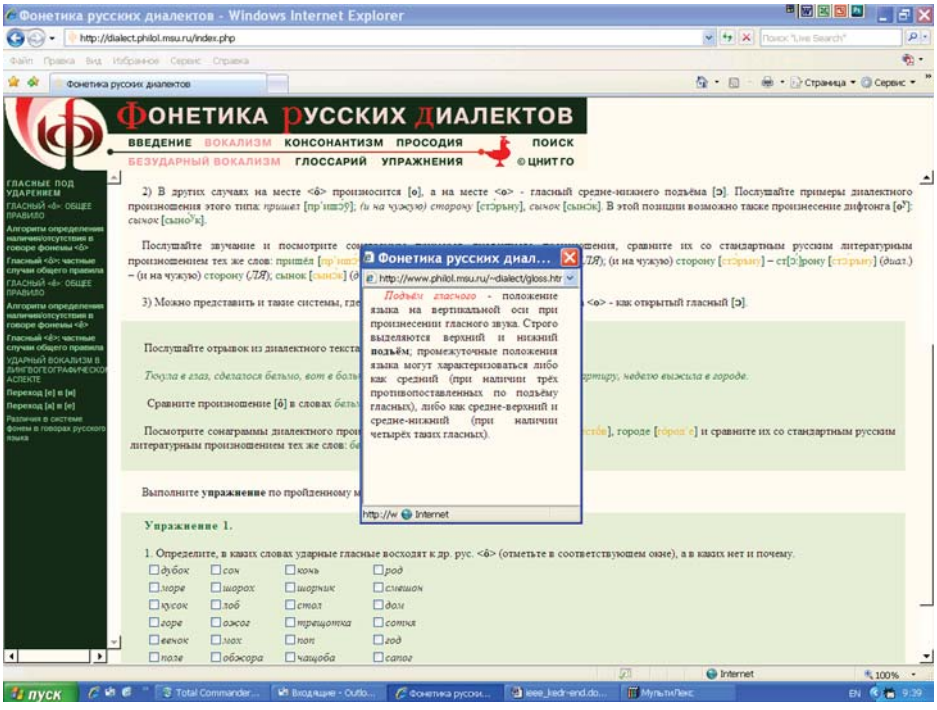


Рис. 7. Окно терминологических дефиниций

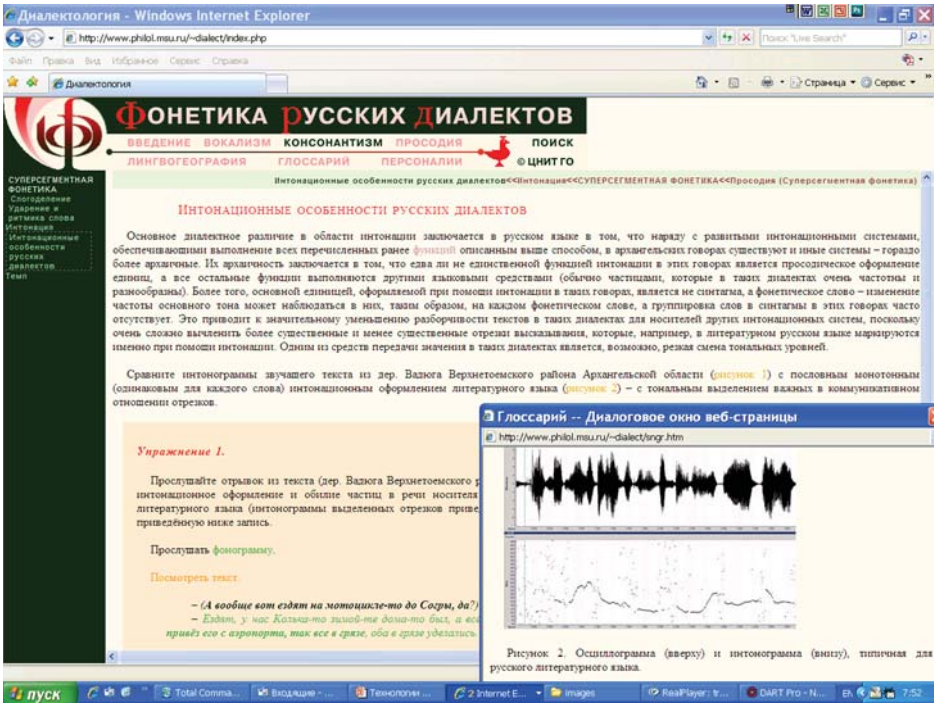


Рис. 8. Окно графического представления акустических характеристик (осциллограмма и интонограмма)

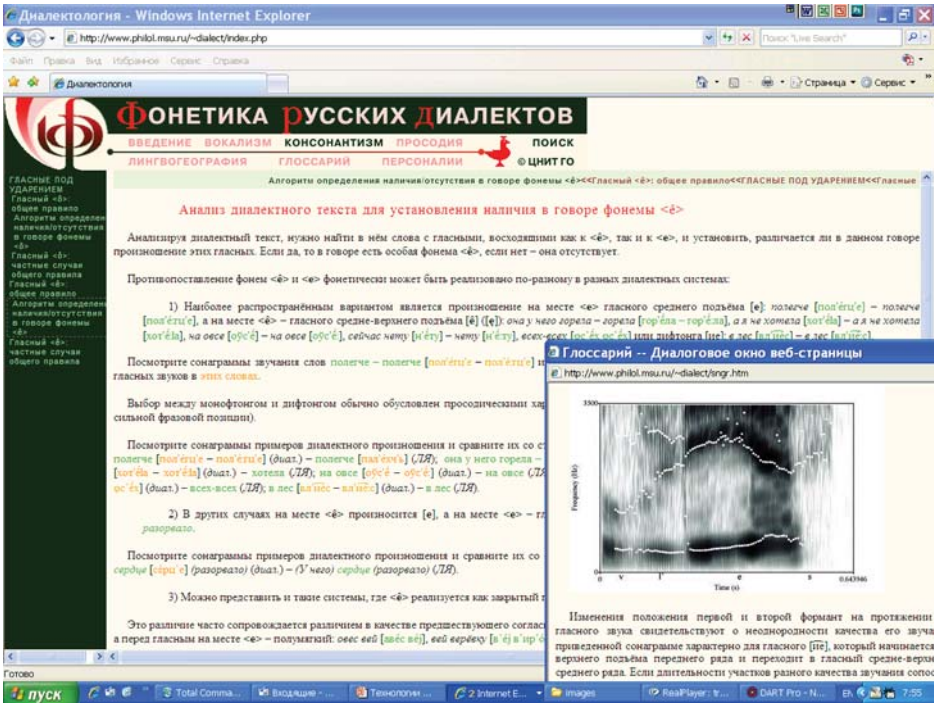


Рис. 9. Окно графического представления акустических характеристик (сонограмма и её описание)

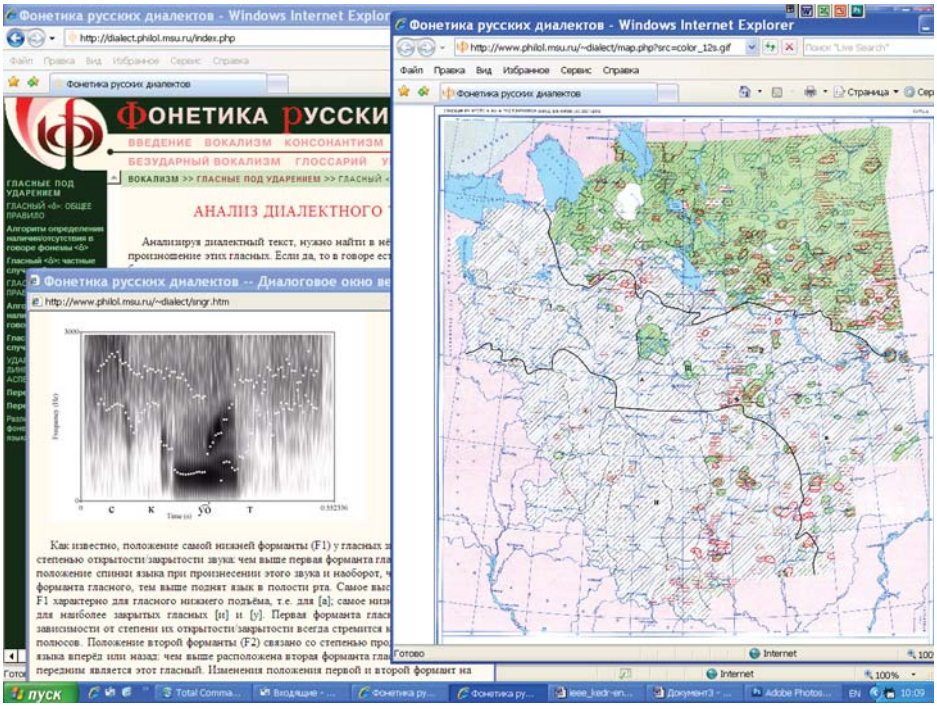


Рис. 10. Окна графических иллюстраций (сонограмма и её описание, диалектологическая карта)

- персоналии (краткие очерки научных интересов и открытий российских филологов, которые внесли значительный вклад в теоретическую и экспериментальную фонетику);
- терминологический словарь (глоссарий терминов и терминологических понятий, используемых в гиперпространстве учебно-справочного сайта).

Таким образом, впервые в пространстве Рунета у всех интересующихся и занимающихся теоретической и практической фонетикой русского языка появилась возможность не только прочитать информацию об особенностях фонетической и интонационной систем, но и увидеть последовательность артикуляторных движений, определяющих русскую произносительную базу, услышать реальное звучание речи на русском языке в режиме on-line.

В разделе, посвящённом сегментной фонетике, подробно разбираются артикуляторные, акустические и перцептивные корреляты русских звуков (системы вокализма и консонантизма), звуки языка и речи, понятие слога и коартикуляторные процессы, понятие фонетического слова, или ритмической структуры, в применении к русскому языку, редукция, ассимилятивные и диссимилятивные процессы, фонетические процессы на стыке слогов и слов и в консонантных сочетаниях и т.п. — т.е. все те явления русской звучащей речи, которые вызывают наибольшие затруднения у иностранных учащихся. Интонационная характеристика высказывания включает в себя описание как дифференциальных, так и интегральных признаков. Рассматриваются основные функции интонации: коммуникативная, выделительная, организующая и эмоциональная. Особое внимание в Интернет-учебнике уделено поддержке правильного (нормативного) русского произношения, для чего внесён специальный отдельный блок материалов по современной русской орфоэпии.

Разработанные авторами в ходе выполнения проекта по проектированию и наполнению Интернет-портала по фонетике русского языка методики построения учебного гипертекста доказали свою эффективность и могут быть рекомендованы в качестве методических указаний для самого широкого круга разработчиков аналогичных учебных материалов, предназначенных для размещения в Сети.

Описываемые учебно-справочные и научно-образовательные Интернет-ресурсы по русской фонетике и фонетике русских диалектов не имеют аналогов ни в России, ни за рубежом, так как являются уникальными по степени проработанности и объёму представленного материала, что подтверждено многочисленными отзывами, приходящими в адрес авторов по электронной почте, и результатами его обсуждения на международных и российских конференциях и семинарах.

Литература

[1] Садовничий В.А., Угольников А.Б., Варламов В.В., Воеводин Вл.В., Кедрова Г.Е., Сергиевская А.Л. От сети профессионалов к профессиональной сети науки и образования России: научно-образовательные ресурсы Московского университета в Интернет // «Телематика-2002». Труды Всероссийской научно-методической конф. С-Петербург, 2002.

[2] Кедрова Г.Е. Системные требования к проектированию электронных учебных материалов для дополнительного профессионального образования // Материалы межрегиональной университетской научно-практической конф. «Современное состояние, проблемы и перспективы развития дополнительного профессионального образования в российских регионах». РГГУ. М.: Каллиграф, 2006.

[3] P. Whalley. An alternative Rhetoric for Hypertext. In.: C. McKnight, A. Dillon & J. Richardson (eds). Hypertext — a psychological perspective. HUSAT Research Institute. 1993.

[4] Егоров А.М., Кедрова Г.Е. Программа обработки компьютерных словарей для исследовательских и учебных целей // Теория и практика речевых исследований (АРСО-99). Материалы конференции. М., 1999.

[5] Болла К. Атлас звуков русской речи. Будапешт, 1981.

Г.Е. Кедрова,
кандидат филологических наук, доцент,
МГУ им. М.В. Ломоносова, филологический факультет.
В.В. Потапов,
доктор филологических наук,
старший научный сотрудник
МГУ им. М.В. Ломоносова, филологический факультет.
А.М. Егоров,
научный сотрудник
МГУ им. М.В. Ломоносова, филологический факультет.
Е.Б. Омельянова,
младший научный сотрудник
МГУ им. М.В. Ломоносова, филологический факультет.
М.В. Волкова,
инженер,
МГУ им. М.В. Ломоносова, филологический факультет.

Опыт использования компьютера при исследовании и тренировке слухо-речевого восприятия у пациентов после кохлеарной имплантации

В.В. Люблинская,
кандидат биологических наук

Е.А. Огородникова,
кандидат биологических наук

И.В. Королёва,
доктор психологических наук

С.П. Пак,
кандидат биологических наук

М.В. Рыбаков

В работе описывается система на базе персонального компьютера, разработанная для восстановления слухоречевой функции у глухих людей после операции кохлеарной имплантации, используемая в процессе реабилитации пациентов. Исследование проводилось совместно сотрудниками лаборатории психофизиологии речи Института физиологии им. И.П. Павлова РАН и Научно-исследовательского института уха, горла, носа и речи (Санкт-Петербург). Основные усилия авторов были направлены на поиск адекватных способов развития базовых навыков слухового восприятия — распознавания речевых сигналов, определения голосовых характеристик говорящего и фразовой интонации, различения звуков окружающего мира и акустической ориентации.

Введение

В современных условиях развития вычислительной техники одним из актуальных направлений является использование программных средств для помощи в организации процесса диагностики и лече-

ния людей с нарушениями различных функций. Вариант подобной системы для повышения эффективности курса реабилитации слухоречевого восприятия у глухих пациентов после операции кохлеарной имплантации в течение ряда лет разрабатывается специалистами Института физиологии им. И.П. Павлова и НИИ уха, горла, носа и речи. В основу системы заложены как данные экспериментального исследования механизмов восприятия речи, так и 10-летний опыт клинической работы с пациентами на базе НИИ ЛОР — одного из пионеров внедрения практики кохлеарной имплантации в России.

Что такое кохлеарная имплантация

Кохлеарная имплантация (электродное протезирование слуха) представляет собой наиболее современный и эффективный метод реабилитации глухих людей. Применение этого метода переживает в настоящее время бурный рост как во всём мире, так и на территории России. Литература по этой тематике огромна, особенно англоязычная (например, см. обзор [1]). Но существуют также хорошие отечественные описания сущности и методов кохлеарной имплантации [2, 3].

В общем виде кохлеарное протезирование включает два этапа. Первый — хирургическая операция по введению во внутреннее ухо (улитку) пациента ряда электродов, обеспечивающих электрическую стимуляцию сохранившихся волокон слухового нерва. Второй — послеоперационная реабилитация, направленная на развитие и восстановление слухоречевой функции пациента [4]. Этот этап является длительным (может занимать несколько лет) и чрезвычайно важным. От его правильной организации в значительной степени зависят общие результаты протезирования. Одним из приоритетов здесь выступают специальные занятия с сурдопедагогом, направленные на формирование у пациентов новых звуковых образов речевых и неречевых сигналов. Необходимость подобного обучения объясняется тем, что характеристики передачи акустической информации кохлеарным имплантом в нейронные отделы существенно отличаются от соответствующей обработки звукового сигнала в слуховой системе человека. Поэтому от сурдопедагога требуется многократное произнесение слов, фраз, отдельных фонем или записей различных звуков для закрепления в памяти пациента «новых» звуковых образов и развития у него способности различения их на слух. Особенно важно и эффективно проведение электродного протезирования у детей раннего возраста [5, 6].

Таким образом, восстановительно-корректирующий курс оказывается достаточно трудоёмким процессом. Для его методического обеспечения могут эффективно использоваться современные средства вычислительной техники и компьютерные программы. Специалисты-дефектологи дают убедительные обоснования эффективности использования компьютерной техники в практике обучения детей с патологией слуха и речи (Кукушкина, 1994). Следует подчеркнуть, что принципиальной особенностью концепции тренажёрной системы в отношении к реабилитации пациентов с КИ является её направленность на тренировку слухового восприятия широ-

кого круга звуков окружающего мира, а не только помощь в развитии слухоречевого восприятия.

Компьютерные тренажёры для пациентов с кохлеарными имплантами могут производить подбор, запись и направленное преобразование речевых (и других звуковых) сигналов, сопровождать их визуальным подкреплением (картинки, надписи), а также формировать различные фоновые условия прослушивания. Такой подход открывает новые возможности для коррекции слуха и позволяет включить в систему реабилитации дополнительные направления, связанные с тренингом помехоустойчивости и инвариантности слухового восприятия, пространственной ориентации и анализа сложной акустической сцены. Важным компонентом выступает также обеспечение возможности объективной оценки динамики показателей обучения пациента.

Общее описание системы и подхода к организации процедуры тренинга

Разработанная система [6, 7] направлена на развитие и восстановление слухоречевой функции как у детей (начиная с 4–5 лет), так и у взрослых пациентов с КИ. Она адаптирована для носителей русского языка и может рассматриваться как дополнительное инструментальное средство для организации процесса обучения под руководством сурдопедагога или самостоятельных занятий пациента. Важная функция системы состоит в возможности объективного контроля за результатами проведённого тренинга, динамикой развития и закрепления перцептивных навыков, что позволяет целенаправленно корректировать содержание индивидуального курса занятий с пациентом в ходе его реабилитации.

Система реализована на основе стандартного вычислительного комплекса — стационарный или портативный компьютер с выносными колонками (динамиками), системой WINDOWS (WIN'98 и выше) и офисным пакетом (Microsoft Office). Типичные элементы системы представлены на рис. 1.

Рис. 1. Общий вид системы

Специальная управляющая программа (программист Рыбаков М.В.) представляет основу программного обеспечения системы и обеспечивает порядок предъявления тестовых звуков, визуальное подкрепление их на экране монитора для выбора ответов, общий контроль за процедурой тренинга, включая условие стимуляции и регистрации результатов.

Визуальная информация на экране монитора может быть представлена соответствующими «картинками» и письменным текстом, а также элементами обратной связи — положительной (смайлик и красный столбик диаграммы) или отрицательной (отсутствие смайлика и серый столбик диаграммы). Примеры видеоизображений экрана для пользователя программы приведены на рис. 2.

Программа позволяет производить следующие действия в процессе работы:
выбрать режим работы (обучение или тестирование),

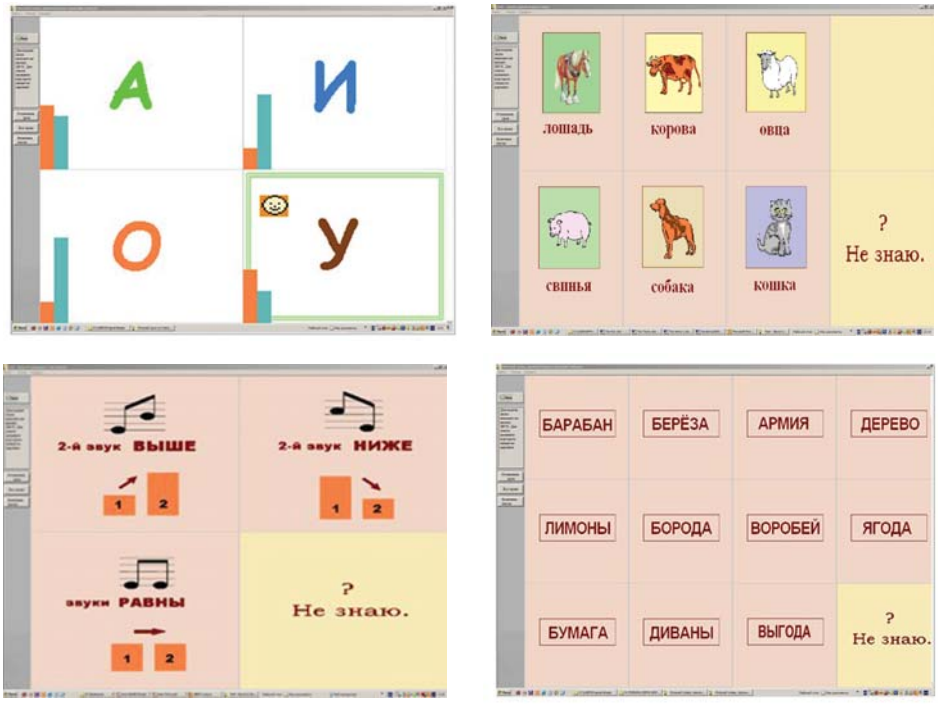


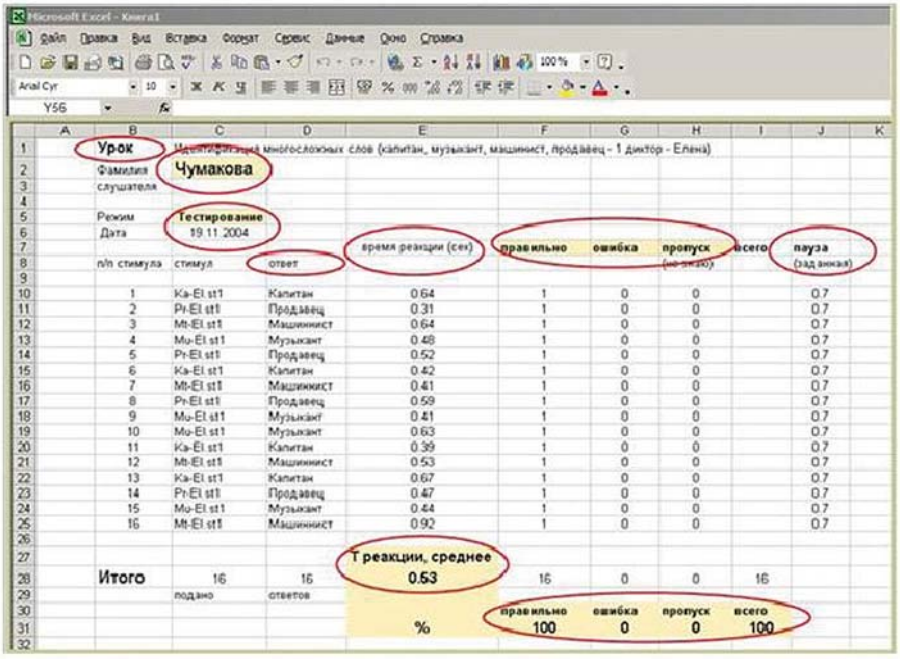
Рис. 2. Примеры ответов, изображаемых на экране монитора при разных типах тестов. Сверху: слева – распознавание изолированных гласных; справа – узнавание характерных звуков животных. Снизу: слева – различение высоты музыкальных звуков; справа – распознавание слов в условиях «конкуренции»

выбрать направление тренинга (набор упражнений),
выбрать порядок стимуляции (произвольный запуск звукового сигнала или последовательность стимулов с заданной паузой между ними),
зафиксировать полученные ответы с помощью протокола занятий в памяти компьютера.

Итоговый протокол тренинга содержит информацию о пациенте (Ф.И.О.), дату и время проведения занятия, обозначение конкретного упражнения, список подаваемых сигналов и соответствующих им ответов, первичные результаты анализа — число и процент правильных опознаний и ошибок, пропусков ответа, среднее время реакции. Вся информация представлена в форме таблиц (программа EXCEL), которые удобны для дальнейшей обработки индивидуальных данных, а также сравнения результатов тренинга по всей группе пациентов. Пример одной из таких таблиц приведён на рис. 3.

Занятия по развитию и закреплению слухоречевых навыков производятся в режиме «ОБУЧЕНИЕ». Он предусматривает не только сопровождение акустической стимуляции визуальным отображением на экране монитора (картинки, надписи), но и возможность повторного прослушивания звуков и включение обратной связи с пациентом.

Обратная связь представлена вариантами как положительного (увеличение красного столбика диаграммы, появление улыбающегося смайлика), так и отрицательного подкрепления (отсутствие смайлика и рост серого стол-



Microsoft Excel - Книга1

Файл Правка Вид Вставка Формат Сервис Данные Окно Справка

	A	B	C	D	E	F	G	H	I	J	K
1		Урок	Идентификация многосложных слов (капитан, музыкант, машинист, продавец - 1 диктор - Елена)								
2		Фамилия	Чумакова								
3		слушателя									
4											
5		Режим	Тестирование								
6		Дата	19.11.2004								
7					время реакции (сек)	правильно	ошибка	пропуск	всего	пауза	
8		п/п стимула	стимул	ответ						(задержка)	
9											
10		1	Ka-El st1	Капитан	0.64	1	0	0		0.7	
11		2	Pr-El st1	Продавец	0.31	1	0	0		0.7	
12		3	Mt-El st1	Машинист	0.64	1	0	0		0.7	
13		4	Mu-El st1	Музыкант	0.48	1	0	0		0.7	
14		5	Pr-El st1	Продавец	0.52	1	0	0		0.7	
15		6	Ka-El st1	Капитан	0.42	1	0	0		0.7	
16		7	Mt-El st1	Машинист	0.41	1	0	0		0.7	
17		8	Pr-El st1	Продавец	0.59	1	0	0		0.7	
18		9	Mu-El st1	Музыкант	0.41	1	0	0		0.7	
19		10	Mu-El st1	Музыкант	0.63	1	0	0		0.7	
20		11	Ka-El st1	Капитан	0.39	1	0	0		0.7	
21		12	Mt-El st1	Машинист	0.53	1	0	0		0.7	
22		13	Ka-El st1	Капитан	0.67	1	0	0		0.7	
23		14	Pr-El st1	Продавец	0.47	1	0	0		0.7	
24		15	Mu-El st1	Музыкант	0.44	1	0	0		0.7	
25		16	Mt-El st1	Машинист	0.92	1	0	0		0.7	
26											
27					T реакции, среднее						
28		Итого	16	16	0.53	16	0	0	16		
29			подано	ответов							
30						правильно	ошибка	пропуск	всего		
31					%	100	0	0	100		
32											

Рис. 3. Пример протокола одной сессии работы с системой

бика диаграммы). Пример ответов с обратной связью приведён на верхнем левом кадре рис. 2. Выбор варианта ответа на экране монитора осуществляется с помощью щелчка «мышкой» по соответствующей ему картинке или надписи на экране монитора.

Проверка и оценка проведённого обучения производится в режиме «ТЕСТИРОВАНИЕ». В этом случае прослушивание звуковых сигналов организовано без повторов и обратной связи. Пациент видит только подтверждение своего ответа — выделение контурной рамкой выбранного варианта ответа.

По завершении каждого занятия (независимо от режима) на экране появляется дополнительное всплывающее окно, содержащее экспресс-оценку проведённого урока, в которую входят число правильных ответов, число ошибок, среднее время реакции и общая отметка за урок. Положительная отметка (знак «+») соответствует уровню, превышающему 70% правильных слуховых опознаний. В противном случае появляется отрицательная отметка (знак «–»), показывающая, что уровень правильных опознаний оказался ниже 70%. Такая информация помогает пациенту оценить свои успехи или неудачи, а также способствует повышению мотивации к дальнейшим занятиям и улучшению выделенных показателей.

Оформление экспресс-оценки производится системой автоматически и не требует запроса пользователя или сурдопедагога. Однако она не фиксируется в памяти компьютера. Для сохранения полученных результатов в полном объёме требуется выбрать опцию «сохранить протокол занятий» и ввести необходимые данные для обозначения пациента (фамилия, код или др.).

Ранее уже отмечалось, что оформление протокола соответствует таблицам в формате специального пакета для статистического анализа данных (EXCEL). При этом первичный

уровень обработки (вычисление средних значений для правильных ответов, ошибок и времени реакции) производится автоматически.

Полученные таблицы сохраняются на жёстком диске (или других носителях памяти) и могут быть выведены на экран (или печать) сразу по окончании занятий.

Разделы и направления тренинга

Структура системы представлена рядом самостоятельных блоков. Такое построение позволяет дополнять и расширять диапазон тренируемых навыков и звукоречевую базу занятий. В настоящее время пользователю доступны пять основных направлений тренинга, которые условно обозначены — «РЕЧЬ», «ПРОСОДИКА», «ОРИЕНТАЦИЯ», «ЗВУКИ ВОКРУГ НАС» и «ВОСПРИЯТИЕ НА ФОНЕ ПОМЕХ».

Целью каждого из разделов является формирование и развитие определённого навыка слухоречевого восприятия — распознавания речевых сигналов (от изолированных фонем до многосложных слов), просодических характеристик речи (голос, интонация), акустической ориентации, различения звуков окружающего мира (включая музыкальные инструменты), выделения целевого сигнала в условиях фоновой помехи или «конкуренции» («речевой коктейль»).

Каждый из разделов включает набор отдельных занятий (уроков), который также может пополняться по мере развития системы. При проведении ряда занятий предусмотрена возможность работы с полным или с сокращённым списком звуковых стимулов, что позволяет выбрать адекватный для пациента уровень сложности перцептивного задания.

Специализированные разделы системы и направления тренинга представлены следующим образом.

В раздел «РЕЧЬ» входит набор уроков, связанных с распознаванием речевых сигналов — изолированных гласных, одно-, разно- и многосложных слов. Он представлен в достаточно ограниченном объёме — от 3 до 5 различных стимулов для каждого из уроков. Основным моментом, направленным на методическую помощь в слуховой работе с пациентом, здесь выступает дикторская вариативность. Система даёт возможность прослушивания речевых сигналов в исполнении разных дикторов. В настоящее время банк дикторов соответствует 4 вариантам голосов (2 мужских и 2 женских), которые перекрывают частотный диапазон голосов взрослых людей (основной тон — от 100 до 250 Гц). Таким образом, тренинг направлен на обучение инвариантному распознаванию речи независимо от голосовых характеристик говорящего.

Раздел «ПРОСОДИКА», напротив, направлен на обучение навыкам выделения и оценивания изменений голосовых характеристик в речи. Раздел представлен набором уроков, связанных с умением различать голоса дикторов (женский-мужской) и интонацию высказывания (утверждение, вопрос) в

соответствии с изменением контура основного тона («Это барабан?» или «Это — барабан.»). Занятия по различению голоса диктора проводятся на всём речевом материале (от гласных до многосложных слов, 4 диктора), занятия по различению интонации — на материале 12 коротких фраз (2 диктора — мужчина и женщина).

Следующий раздел тренинга («ЗВУКИ ВОКРУГ НАС») представлен достаточно традиционным набором разнообразных звуков окружающей среды: голосами животных, птиц, звуками дома, улицы, стихии, музыкальными сигналами (звучание музыкальных инструментов). Этот раздел ориентирован, в основном, на рано оглохших пациентов и предусматривает возможность выбора занятий с полным (более 6 категорий) и сокращённым (до 5 вариантов) списком стимулов.

В настоящее время разрабатывается дополнительный раздел «музыкального» тренинга, включающий дополнительные возможности по развитию звуковысотного восприятия и ритмики.

Разделы «ОРИЕНТАЦИЯ» и «ВОСПРИЯТИЕ НА ФОНЕ ПОМЕХ» представляют наиболее сложные задачи тренинга (акустическая ориентация в пространстве и помехоустойчивость слухоречевого восприятия). Однако в то же время они хорошо демонстрируют методические возможности системы, которые связаны с элементами моделирования сложной перцептивной среды.

Так, в разделе «ОРИЕНТАЦИЯ» представлен набор уроков, помогающий сформировать начальные навыки пространственной ориентации. Это особенно актуально, учитывая односторонний характер операции имплантации и объективные ограничения возможности пространственного восприятия у пациентов. Уроки включают обучение различению пространственного положения (локализации) источника звука (посылки шума) или речи (разносложные слова, 4 диктора). При этом можно использовать варианты латеральной (справа–слева) и фронтальной (впереди–сзади или дальше–ближе) схемы расположения динамиков. Однако в период начальной реабилитации наиболее целесообразно использование латерального варианта размещения, максимально ориентированного на восприятие монауральных признаков локализации.

Следующие наборы уроков в данном разделе связаны с тренингом навыков обнаружения движения источника звука (источник стоит или движется) и определения его направления (движется слева направо или справа налево). Эти перцептивные задания представляют для пациентов серьёзную трудность, но чрезвычайно актуальны для их дальнейшей слуховой практики. В раздел входит также дополнительное занятие по проверке слухового распознавания речевых сигналов в условиях изменения пространственной позиции диктора (разносложные слова, 4 диктора).

В раздел «ВОСПРИЯТИЕ НА ФОНЕ ПОМЕХ» входит два основных набора упражнений. Первый из них прямо соответствует названию раздела и связан с тренингом выделения и опознания речевого сигнала на фоне различных акустических помех — шума, речи и музыки. В качестве речевого сигнала здесь также выступают разносложные слова в исполнении четырёх дикторов. Помехой являются фрагменты «белого» шума, текста (мужской голос), инструментальной и вокальной музыки. Интенсивности сигнала и помехи выравнены 1:1.

Второй набор соответствует наиболее сложным условиям восприятия, моделирующим известный эффект «cocktail party». При выполнении задания от пациента требуется

выделить и опознать целевой стимул (например, изолированный гласный звук или многосложное слово, сказанное женским голосом) в условиях прямой конкуренции — одновременное произнесение речевых сигналов разными дикторами. Выполнение этих заданий вызывает затруднения даже у слушателей с нормальным слухом (до 20–25% ошибок).

Апробация системы

Система в течение ряда лет успешно используется в Институте уха, горла, носа и речи при реабилитации пациентов с КИ различного возраста и уровня языковой компетенции [8].

Опыт работы (более 40 пациентов) свидетельствует, что и у детей, и у взрослых пациентов наблюдается высокий уровень мотивации к занятиям с применением компьютера и специальной программы. Важным моментом, подтверждённым на практике, является возможность адаптации занятий, проводимых с помощью системы, к разным этапам курса реабилитации и степени индивидуального развития навыков слухоречевого восприятия пациента.

Результаты апробации системы показали также, что она удобна как для организации слухового тренинга под руководством сурдопедагога/родителя, так и для самостоятельной работы ребёнка/взрослого, в том числе на домашнем компьютере. Кроме того, система отвечает методическим требованиям и может быть использована и для проведения целевых научных исследований, результаты которых опубликованы в ряде работ, посвящённых восприятию высоты [9], разделению звуковых потоков на примере конкурирующих гласных [10], узнаванию звуков музыкальных инструментов [11], возможности ориентации в звуковом пространстве [12].

Приведём конкретный пример результатов одного теста из раздела «ВОСПРИЯТИЕ НА ФОНЕ ПОМЕХ»: распознавание речевых сигналов в условиях эффекта «cocktail party», когда предъявляемые стимулы представляют собой гласные или слова, произнесённые одновременно двумя дикторами. Это один из сложных тестов, выполнение которого вызывает затруднения даже у нормально слышащих испытуемых. Так, для первой части раздела — тесты с конкурирующими гласными звуками — процент ошибочных распознаваний обоих звуков слушателями с нормой слуха может достигать 25%. Большинство пациентов с КИ вообще не способны правильно опознавать оба одновременно звучащих гласных, достаточно хорошо они могут идентифицировать только один из них — преимущественно гласный, произнесённый мужским голосом. Более подробно описание процедуры и результаты этого теста опубликованы в статье [10].

Во второй части раздела использовался тест, где в качестве стимулов пациентам предъявлялись изолированные слова, произнесённые одновременно двумя дикторами — мужчиной и женщиной. Процедура тестирования состояла из двух сессий. В одной из них от испытуемого требовалось опознать слова, произнесённые женским голосом, в другой — опознать слова,

произнесённые мужским голосом. В обоих случаях предъявлялся один и тот же набор стимулов: Ягода + *Армия,* Борода + *Барабан,* Воробей + *Берёза,* Бумага + *Дерево,* Диваны + *Лимоны,* Ягода + *Выгода* (прямым шрифтом написаны слова мужского голоса, курсивом — женского.)

Стимулы в случайном порядке предъявлялись слушателю (опыт выполнялся с каждым испытуемым индивидуально). Одновременно на мониторе компьютера изображалась таблица с надписями всех слов, из которых составлены стимулы. Слово — ответ испытуемый выбирал курсором «мышки», он записывался в компьютере в специальный файл в текстовом формате с помощью программы EXCEL.

Было обследовано 6 пациентов — взрослых, оглохших поздно, до глухоты владеющих речью. 5 пациентов проходили тестирование по одному разу в режиме обучения. Рис. 4, иллюстрирующий их результаты, показывает довольно ограниченную способность опознавания слов целевого диктора: процент правильных ответов здесь составляет, в

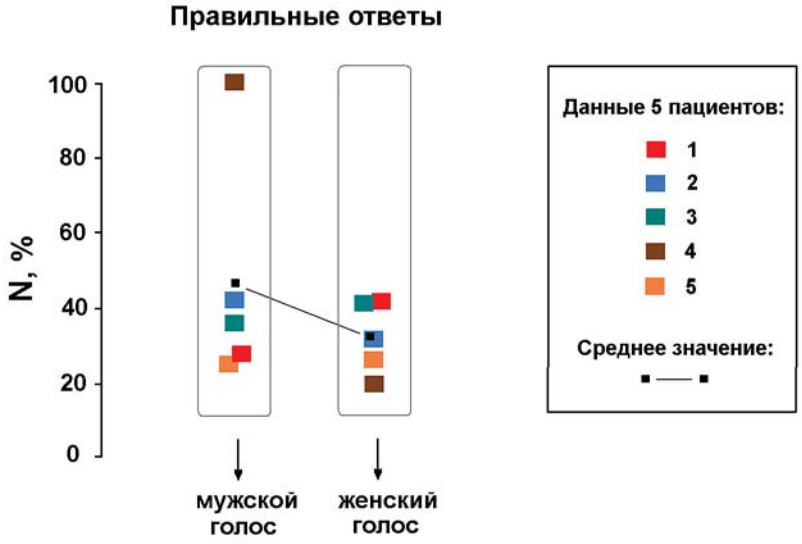


Рис. 4. Результаты тестирования пациентов при опознавании слов, произнесённых одновременно мужским и женским голосом (пояснения в тексте)

среднем, 46% для слов, сказанных мужским голосом, и 32.2% для слов, произнесённых диктором-женщиной. Исключение составляет только 100% опознаний слов мужского голоса у пациента №4. Но для женского голоса его результат не превышает 20% правильных ответов. Отметим, что этот пациент выделяется из группы и по опыту использования КИ — более двух лет.

Один из пациентов прошёл тесты в режиме обучения последовательно три раза с некоторым интервалом. Его данные, приведённые на рис. 5, иллюстрируют возможность тренировки способности к раздельному распознаванию смешанных слов: при последовательных занятиях показатели (% правильных ответов и время реакции) заметно улучшались раз от разу.

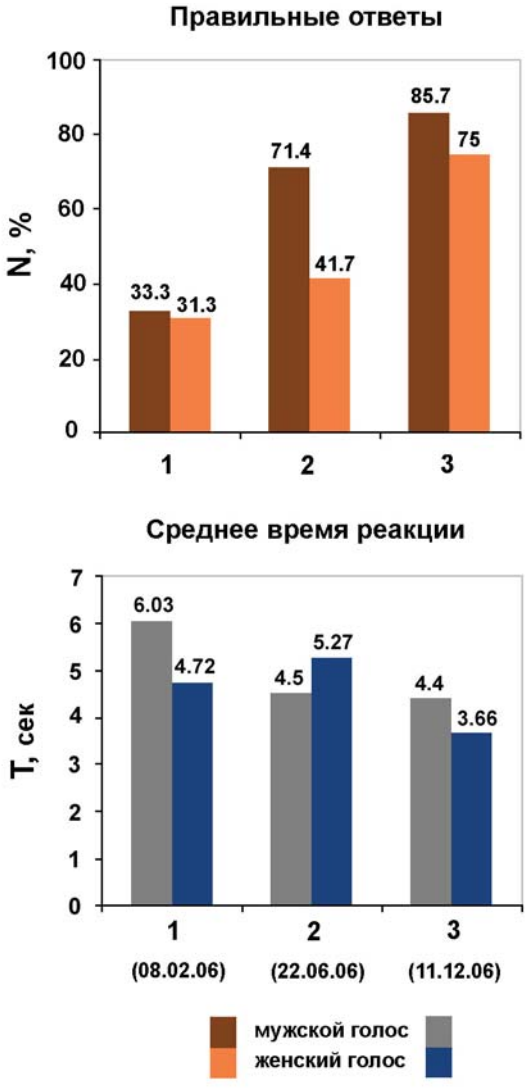


Рис. 5. Результаты одного пациента (№ 6) при последовательном тестировании в разное время. Верхний график — % ответов правильного опознавания, нижний график — время реакции (в сек). Подписи внизу — дата проведения тестов

Заключение

В целом, применение системы с набором тематических разделов тренинга и объективной оценкой обучения повышает эффективность реабилитационного процесса у пациентов, способствует сокращению сроков их социальной адаптации, а также нагрузки на специалиста-сурдопедагога, ответственного за слуховую работу с пациентом.

Представленная система поддержана рядом патентов РФ, получила положительную оценку специалистов в области сурдопедагогики (РГПУ им. А.И. Герцена) и была успешно представлена на Выставке инновационных достижений России в рамках XI Международного экономического форума (Санкт-Петербург, 2007) [13, 14].

Литература

1. Clark G. Cochlear Implants. Speech Processing in the Auditory System. Eds.: Greenberg S., Ainsworth W.A., Popper A.N., Fay R.R. «Springer». 2004. P.422–462.

2. Кохлеарная имплантация. Учебное пособие / Составитель Таварткиладзе Г.А. М., 2000. 81 с.

3. Королёва И.В. Слухоречевая реабилитация глухих детей и взрослых с кохлеарными имплантами. СПб.: ЛЕМА, 2007. 104 с.

4. Королёва И.В. Развитие слухоречевого восприятия после кохлеарной имплантации у глухих школьников и взрослых: Учебное пособие. СПб.: С.-Петербургский НИИ уха, горла, носа и речи, 2008. 200 с.

5. Королёва И.В. Этапы развития слухоречевого восприятия и речи у рано оглохших детей с кохлеарным имплантом // Рос. оторинолар. 2008. №1. С. 11–20.

6. Королёва И.В. Слухоречевая реабилитация глухих детей с кохлеарными имплантами. СПб: НИИ уха, горла, носа и речи, 2005, 90 с.

7. Кукушкина О.И. Компьютер в специальном обучении. Проблемы, поиски, подходы. Дефектология, №5, 1994. С.3–10.

8. Огородникова Е.А., Королёва И.В., Люблинская В.В., Пак С.П., Столярова Э.И. Использование компьютерных программ в процессе слухоречевой реабилитации пациентов с кохлеарными имплантами / Актуальные вопросы логопатологии / Под ред. И.В. Королёвой: Сб.ст. СПб. НИИ уха, горла, носа и речи. СПб., 2004. С. 73–77.

9. Огородникова Е.А., Королёва И.В., Люблинская В.В., Пак С.П. Компьютерная тренажёрная система для реабилитации слухоречевого восприятия у пациентов после операции кохлеарной имплантации // Рос. оторинолар. Приложение №1, 2008. С. 342–347.

10. Королёва И.В., Огородникова Е.А., Люблинская В.В., Пак С.П., Балякова А.А. Результаты использования компьютерной тренажёрной системы в практике реабилитации слухоречевого восприятия у пациентов с кохлеарными имплантами // Рос. оториноларинг. Приложение №1, 2008. С. 297–304.

11. Люблинская В.В., Королёва И.В., Огородникова Е.А., Пак С.П., Столярова Э.И. Восприятие высоты голоса и мелодики речевых сигналов глухими людьми с кохлеарными имплантами // Рос. оториноларинг., 2007, №4. С. 3–13.

12. Люблинская В.В., Королёва И.В. Разделение звуковых потоков глухими людьми с кохлеарным имплантом // Сенсорные системы, 2006. Т.20. №3. С. 195–203.

13. Королёва И.В., Росс Я.Ю., Огородникова Е.А. Восприятие музыкальных стимулов пациентами после операции кохлеарной имплантации // Рос. оториноларинг., 2006, №5. С. 46–54.

14. Огородникова Е.А., Пак С.П., Королёва И.В. Возможности перцептивного тренинга функции акустической ориентации у пациентов с кохлеарными имплантами / Матер. 5-го международного симпозиума «Современные проблемы физиологии и патологии слуха». Суздаль, 2004. С. 141–143.

15. Королёва И.В., Люблинская В.В., Огородникова Е.А., Пак С.П., Столярова Э.И., Пудов В.И. Способ слухоречевой реабилитации и её оценки у пациентов с кохлеарными имплантами / Патент 2209057 Российская Федерация, МПК[7] A 61 F 11/10. Заявитель и патентообладатель СПб НИИ уха, горла, носа и речи, №2002108657/14; заявл. 02.04.02; опубл. 27.07.03, Бюлл. №21. 10 с.

16. Огородникова Е.А., Королёва И.В., Пак С.П. Способ реабилитации функции акустической ориентации и её оценки у пациентов с кохлеарным имплантом / Патент 2265426 Российская Федерация, МПК[7] A 61 F 11/04. Заявитель и патентообладатель СПб НИИ уха, горла, носа и речи, №2004108056/14; заявл. 11.03.04; опубл. 10.12.05, Бюлл. №34. 9 с.

В.В. Люблинская,
кандидат биологических наук, ведущий научный сотрудник Института физиологии им. И.П. Павлова РАН.

Е.А. Огородникова,
кандидат биологических наук, заведующая лабораторией психофизиологии речи Института физиологии им. И.П. Павлова РАН.

И.В. Королёва,
доктор психологических наук, профессор кафедры сурдопедагогики РГПУ им. А.И. Герцена, главный научный сотрудник ГУ ФГУ «Санкт-Петербургский НИИ уха, горла, носа и речи Росмедтехнологий».

С.П. Пак,
кандидат биологических наук, старший научный сотрудник Института физиологии им. И.П. Павлова РАН.

М.В. Рыбаков,
инженер Института физиологии им. И.П. Павлова РАН.

Обеспечение содержательного доступа к информационным ресурсам по компьютерной лингвистике

Ю.А. Загорулько,
кандидат технических наук

Е.Г. Соколова,
кандидат филологических наук

И.С. Кононенко

Г.Б. Загорулько

О.И. Боровикова

В статье рассматривается интернет-портал знаний, обеспечивающий систематизацию и интеграцию знаний и информационных ресурсов по компьютерной лингвистике, а также содержательный доступ к ним (поиск информации и навигацию в терминах предметной области портала). Для целостного представления знаний и информационных ресурсов по компьютерной лингвистике их систематизация и структуризация выполнены на основе онтологии. Благодаря этому вся информация на портале представлена в виде сети взаимосвязанных информационных объектов.
Ключевые слова: портал знаний, компьютерная лингвистика, онтология, информационные ресурсы, содержательный доступ

Введение

В связи с постоянно растущими потребностями в средствах автоматической обработки документов и естественно-языковых, в том числе речевых, интерфейсах возникает необходимость в эффективном доступе не только к публикациям, описывающим методы и подходы, разработанные в лингвистике, но и к разного рода словарям, программным компонентам

и алгоритмам, реализующим различные задачи обработки текста и звучащей речи. В настоящее время в сети Интернет представлен большой объём знаний и информационных ресурсов по этой тематике, однако доступ к ним значительно затруднён, так как они систематизированы лишь частично и к тому же рассредоточены по различным интернет-сайтам, каталогам и электронным архивам.

Для устранения подобных проблем создаются специальные интернет-ресурсы, которые выполняют информационную поддержку разнообразных научных и тематических сообществ. Самым известным ресурсом такого рода, имеющим отношение к компьютерной лингвистике (КЛ), является англоязычный каталог LINGUIST List (http://linguistlist.org/), созданный для общения и обмена знаниями между лингвистами. Он содержит информацию о публикациях, персоналиях, научных учреждениях и других организациях лингвистического направления, грантах, конкурсах, проектах, фондах и источниках финансирования, а также о научных мероприятиях в лингвистической сфере деятельности. Кроме того, LINGUIST List предоставляет возможность поиска ресурсов по таким параметрам, как страна, язык, раздел лингвистики.

Из других зарубежных разработок стоит отметить созданный в Германском Исследовательском Центре Искусственного Интеллекта (DFKI) информационный портал «Language Technology World» (http://www.lt-world.org/). Тематические разделы этого портала содержат информацию о лингвистических технологиях, продуктах и информационных системах в области обработки естественного языка, а также о проектах, организациях, персонах. В основу портала положена онтология, благодаря чему возможно установление связей между его разделами. К сожалению, на этом портале практически отсутствует информация об исследованиях, проводимых в России.

К российским аналогам LINGUIST List можно отнести научно-образовательный портал «Лингвистика в России: ресурсы для исследователей» (http://uisrussia.msu.ru/linguist/index.jsp) и сайт «Российская лингвистика (RUSLING)» (http://rusling.narod.ru), который разрабатывается в Отделении лингвистических исследований ВИНИТИ РАН.

Портал «Лингвистика в России» содержит иерархически организованный каталог ссылок на наиболее значимые лингвистические ресурсы и позволяет осуществлять навигацию по разделам портала с помощью иерархических связей внутри разделов и по ссылкам на связанные с ними области (разделы). Тематические категории данного портала представлены разделами по компьютерной, теоретической и прикладной лингвистике и их приложениям (смежным областям), а также разделами, посвящёнными русскому языку, языкам мира и народов РФ.

Портал «Российская лингвистика» предлагает лингвистам «информационную карту» для поиска информации об организациях, научных исследованиях и публикациях, лингвистических ресурсах и персоналиях. Он содержит обширный каталог ссылок на словари и корпуса текстов для различных языков (в том числе славянских), а также сведения о российских лингвистах, предоставляя возможность их поиска не только по алфавиту, но и по области и объекту (языку) исследования.

Примером специализированного тематического ресурса по КЛ является российский сайт «Речевые технологии» (http://speech-soft.ru/), на котором представлена информация, охватывающая прикладные аспекты данного направления (технологии, программные средства, коллективы разработчиков, конкретные системы и т.п.).

Как правило, научно-практические проекты, разрабатываемые в рамках описанных выше подходов, направлены либо на описание и сохранение общей лингвистической информации, либо на представление информации о каком-то одном разделе лингвистики, но не для интеграции ресурсов по компьютерной лингвистике и обеспечения доступа к ним широкому кругу пользователей.

Для решения этих проблем в рамках предложенного нами подхода к построению специализированных интернет-порталов знаний [1] разработан портал знаний по компьютерной лингвистике. Как информационный ресурс указанный портал обеспечивает следующие возможности:

панорамную характеристику научного направления «компьютерная лингвистика» через представление используемых в нём терминов и понятий, объектов и методов исследования, научных результатов, а также участников научной деятельности в рамках этого направления (персон, групп, сообществ и других организаций, вовлечённых в процесс исследования);

интеграцию доступных информационных ресурсов по компьютерной лингвистике в единое информационное пространство;

содержательный доступ к систематизированным знаниям и данным, относящимся к компьютерной лингвистике, т.е. возможность поиска и получения информации в терминах предметной области портала, а также удобную навигацию по всему информационному пространству портала, базирующуюся на модели его предметной области;

информационную поддержку пользователей, т.е. анонсирование разного рода событий и мероприятий, касающихся данного научного направления.

1. Информационная модель портала знаний по КЛ

В качестве концептуальной основы информационной модели портала знаний выбрана онтология [2], с помощью которой можно достаточно просто обеспечить унифицированное представление и хранение знаний и информационных ресурсов по компьютерной лингвистике, а также содержательный доступ к ним.

Онтология портала обеспечивает представление понятий, необходимых для описания как научной деятельности и научного знания в целом, так и конкретной научной дисциплины в частности. Поэтому онтология портала включает универсальные онтологии научной деятельности и научного знания [3], а также онтологию предметной области.

Универсальные онтологии не зависят от предметной области (ПО) и могут использоваться практически в любом портале научных знаний, независимо от его конкретной тематики. В связи с этим указанные онтологии выделены в качестве базовых (рис. 1). Рассмотрим их подробнее.

Онтология научной деятельности является онтологией верхнего уровня и включает базовые понятия, относящиеся к **организации** научно-

исследовательской деятельности, такие, как *Персона*, *Организация*, *Событие*, *Деятельность*, *Публикация*, которые используются для описания участников научной деятельности, мероприятий, научных программ и проектов, различного типа публикаций, а также *Географическое место*. В эту онтологию также включено понятие *Информационный ресурс*, которое служит для описания информационных ресурсов, представленных в сети Интернет.

Онтология научного знания, по своей сути, является метаонтологией. Она содержит метапонятия и отношения, задающие структуры для описания рассматриваемой предметной области, такие, как *Раздел науки, Предмет исследования, Объект исследования, Метод исследования, Научный результат*, позволяющие выделить в данной науке значимые разделы и подразделы, задать типизацию предметов, объектов и методов исследования, описать результаты научной деятельности.

Понятия базовых онтологий связаны между собой ассоциативными отношениями (см. рис. 1), выбор которых осуществлялся не только исходя из полноты представления моделируемой области знаний портала, но и с учётом удобства навигации по его информационному пространству и поиска информации.

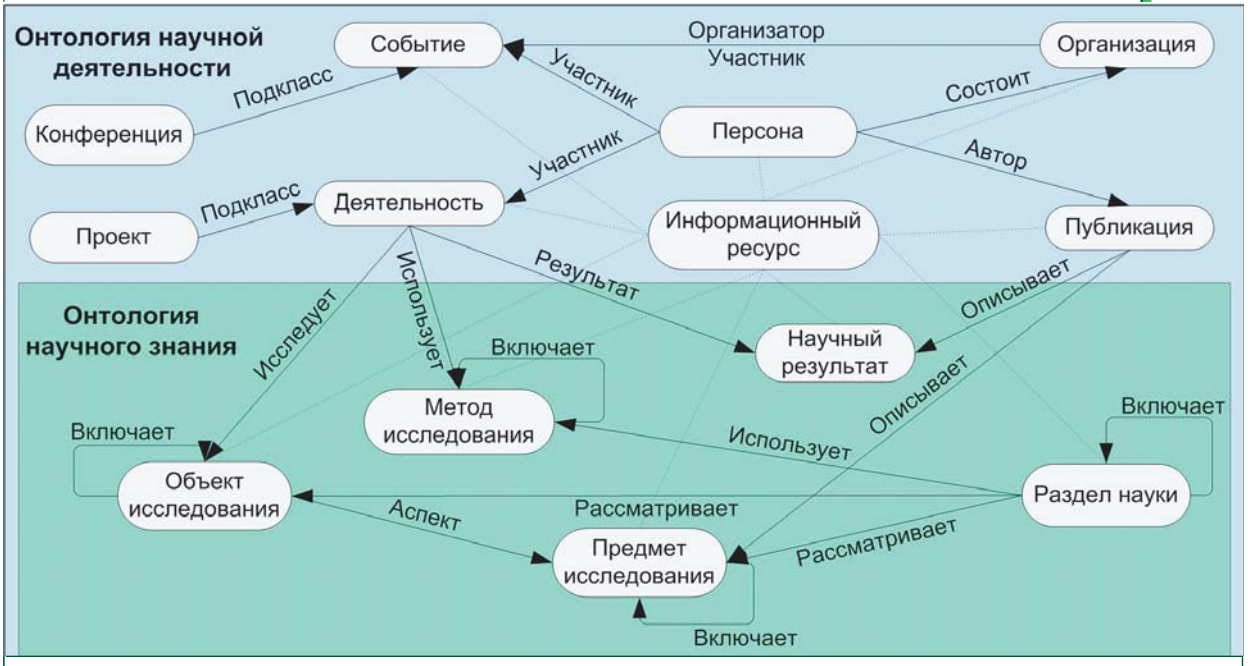


Рис. 1. Базовые онтологии портала

В качестве онтологии предметной области портал включает онтологию компьютерной лингвистики, фрагмент которой представлен на рис. 2. Понятия этой онтологии являются реализациями метапонятий онтологии научного знания и организованы в пять иерархий «общее-частное», каждая из которых соответствует одному из перечисленных выше метапонятий. Все эти иерархии связаны между собой посредством ассоциативных отношений, часть которых наследуется из базовых онтологий, а часть отражает специфику данной предметной области.

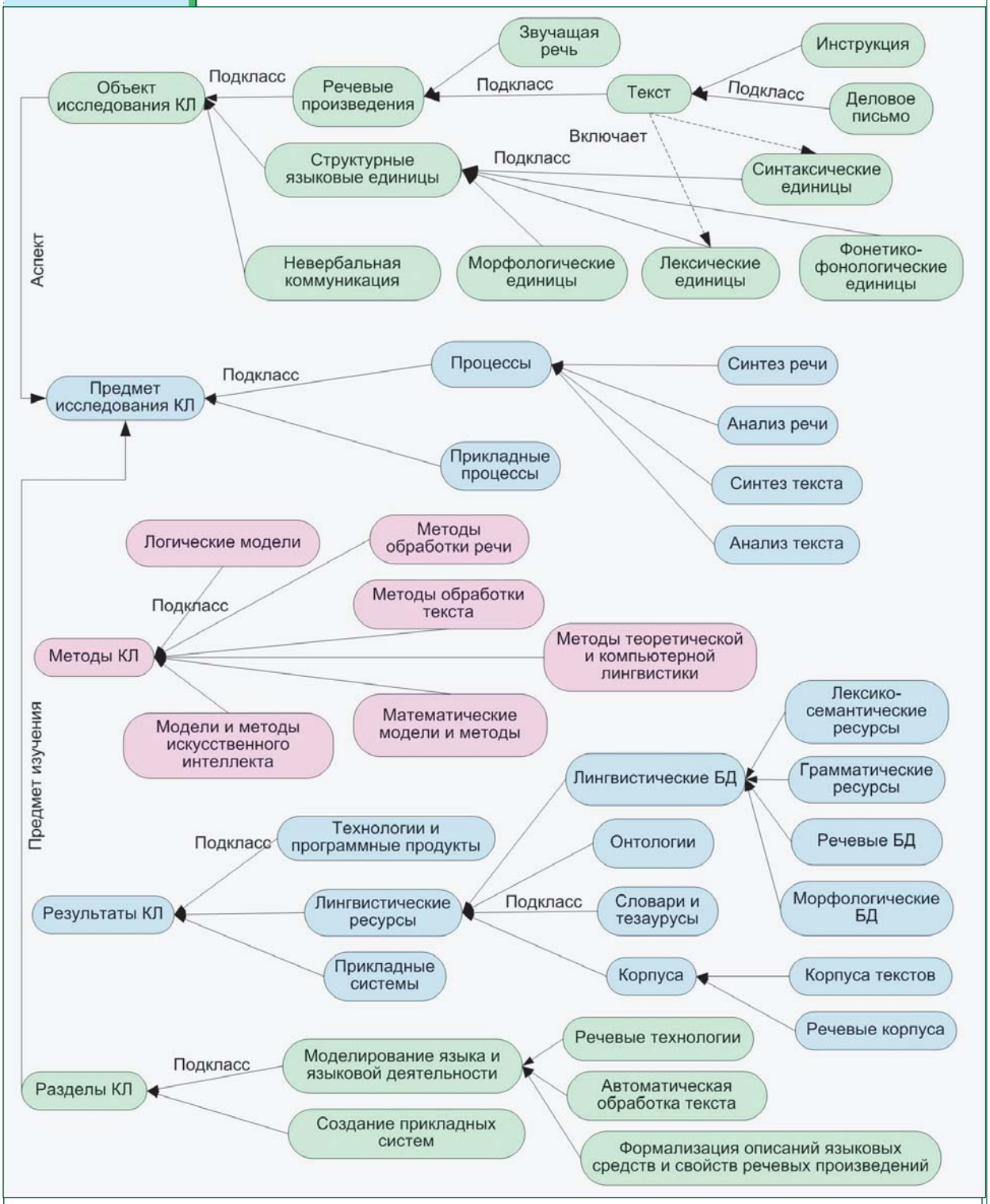


Рис. 2. Фрагмент онтологии компьютерной лингвистики

Рассмотрим онтологию компьютерной лингвистики подробнее.

В качестве **базовых объектов** исследования КЛ предложено рассматривать *Речевое произведение* (РП) как объективную форму существования и использования естественного языка, *Структурные языковые единицы* в составе РП, соответствующие различным языковым уровням: предложения, словосочетания, слова, морфемы, звуки и интонационные единства, а также *Невербальную коммуникацию*.

Класс понятий РП в зависимости от формы (графической или звуковой) представлен в иерархии двумя подклассами: *Текст* и *Звучащая речь*. Выделяемые в РП *Языковые единицы* сгруппированы в соответствии с языковыми уровнями в классы: *Синтаксические единицы, Лексические единицы, Морфологические единицы* и *Фонетико-фонологические единицы*. Для представления связи между целостными РП и их структурными единицами используется отношение «Включение».

Предметом исследования в КЛ являются *Процессы и задачи*, связанные с функционированием языковых единиц в коммуникации, и *Прикладные процессы и задачи*, имеющие практическую ценность, отвечающие определённому социальному запросу. Иерархия предметов исследования связана ассоциативным отношением «Аспект» с иерархией объектов исследования и отношением «Предмет исследования раздела науки» с иерархий разделов науки.

Иерархия методов исследования служит для систематизированного описания инструментов исследования, применяемых в компьютерной лингвистике. В этой иерархии были выделены подклассы понятий *Методы теоретической и компьютерной лингвистики, Методы обработки текста, Методы обработки речи, Модели и методы искусственного интеллекта, Логические модели* и др.

В основе Иерархии разделов КЛ лежит классификация базовых теоретических и прикладных направлений компьютерной лингвистики. В качестве **главных** выделены разделы *Моделирование языка и языковой деятельности* (с подразделами *Автоматическая обработка текста (АОТ), Речевые технологии (РТ), Формализация описаний языковых средств и свойств речевых произведений*) и *Создание прикладных систем*. В зависимости от направления моделирования (анализ или синтез) в классе понятий *Автоматическая обработка текста* выделены соответствующие подклассы: *Понимание текста* и *Генерация текста*, а в классе *Речевые технологии — Распознавание речи* и *Синтез речи*. В зависимости от объекта обработки (текст или звучащая речь), *Прикладные системы* разделены на классы *Создание прикладных систем АОТ* и *Создание прикладных систем РТ*.

Иерархия Научных результатов служит для типизации и описания результатов научной деятельности. В этой иерархии выделены следующие классы: *Технологии и программные продукты, Прикладные системы, Лингвистические ресурсы*. Последний класс делится на такие классы: *Корпуса*, *Лингвистические БД, Онтологии, Словари и тезаурусы*.

Таким образом, вводя формальные описания проблемной и предметной области в виде понятий и отношений между ними, онтология портала задаёт структуры для представления реальных объектов и связей между ними.

В соответствии с принятой нами моделью данные на портале представлены в виде множества разнотипных информационных объектов и связей между ними. *Информационный объект* (ИО) — это структурированная совокупность данных, представляющая собой

описание некоторого объекта из выбранной области знаний или релевантного для неё информационного ресурса. Каждый ИО соответствует некоторому понятию онтологии и имеет заданную им структуру. Между конкретными информационными объектами могут существовать связи, семантика которых определяется отношениями, заданными между соответствующими понятиями онтологии.

2. Информационное содержание портала знаний по КЛ

Информационное содержание (контент) портала включает как знания общего характера, так и конкретные знания о реальных объектах и информационных ресурсах, систематизированные в соответствии с онтологиями портала.

В контенте портала КЛ представлены, прежде всего, знания об основных разделах компьютерной лингвистики, о её предметах и объектах исследования, используемых в ней моделях и методах. Кроме этого, пользователи портала могут найти информацию о выполняемой в области компьютерной лингвистики научной деятельности. В первую очередь, это информация об учёных, исследовательских группах и организациях и их деятельности. Так, например, при просмотре информации о «Группе речевых исследований при кафедре теоретической и прикладной лингвистики филологического факультета МГУ» можно увидеть список исследователей, занятых в деятельности этой группы, а также определить её место в структурной иерархии этого подразделения в рамках университета. Направление работ группы представлено такими разделами КЛ, как *Речевые технологии*, *Создание прикладных систем РТ* и *Формализация описаний языковых средств и свойств речевых произведений*. Кроме того, из описания группы можно перейти на проект ISABASE, в котором она участвовала, и на её сайт, являющийся основным информационным ресурсом этой группы.

В деятельности организаций и исследователей особое место занимают научные и коммерческие проекты, в рамках которых создаются лингвистические знания и ресурсы. Результаты этой деятельности находят отражение в публикациях — монографиях, статьях, материалах конференций и семинаров, отчётах и других текстовых ресурсах, доступ к которым предоставляется порталом. Например, на портале можно найти информацию о монографии Д. Журавского и Дж. Мартина «Speech and Language Processing. An introduction to Natural Language Processing, Computational Linguistics and Speech Recognition», учебнике Р. Миткова «The Oxford handbook of computational linguistics», монографии Н.Н. Леонтьевой «Автоматическое понимание текста: системы, модели, ресурсы» и других публикациях.

Портал обеспечивает доступ и к информационным ресурсам, представляющим непосредственные результаты деятельности организаций и отдельных исследователей, полученные в ходе выполнения научных и коммерческих проектов, а именно: технологии, программные продукты, прикладные системы, лингвистические ресурсы: словари, корпуса (текстов и речи) и лингвистические БД. Для организации более эффективного доступа

к таким ресурсам в контенте представлена информация о различных аспектах их разработки: организациях, персонах и проектах, с которыми связано их появление, а также о таких содержательных характеристиках ресурсов, как отнесённость к разделу науки, объекту или предмету исследования, методам исследования. Эта информация связывает ресурсы с остальными данными и знаниями, представленными в контенте портала, что позволяет пользователю выделить группы ресурсов, созданные, например, в ходе осуществления некоторой исследовательской деятельности (гранта, проекта, конкурса) или с использованием определённого класса методов исследования. Например, при просмотре информации о лингвистическом ресурсе «Речевой корпус RuSpeech» пользователь может заметить, что тематика научного результата объединяет разделы КЛ *Речевые технологии* и *Создание корпусов*. А при дальнейшем переходе к просмотру описания проекта RuSpeech, в рамках которого создавался речевой ресурс, можно увидеть информацию о других результатах и публикациях этого проекта.

Важным компонентом информационного контента портала является описание интернет-ресурсов, к которым относятся сайты организаций, конференций, проектов, порталы и каталоги, а также отдельные страницы с материалами графического, мультимедийного или текстового типа. Как было сказано выше, каждый интернет-ресурс, представленный на портале, соответствует такому понятию онтологии, как Информационный ресурс. Описание отдельного ресурса включает экземпляр данного понятия и набор экземпляров отношений, связывающих его с другими объектами, представляющими организации, персоны, публикации, события и т.д.

3. Обеспечение доступа к ресурсам по компьютерной лингвистике

Основное назначение рассматриваемого портала знаний — обеспечить содержательный доступ к систематизированным знаниям и информационным ресурсам по компьютерной лингвистике. Доступ к знаниям и данным портала осуществляется путём навигации по дереву понятий онтологии и контенту портала (см. рис. 3), а также через развитые средства содержательного поиска.

3.1. Навигация по контенту портала

Для конечного пользователя данные на портале представлены в виде множества связанных информационных объектов. При навигации по порталу обеспечивается возможность выбора ИО, относящихся к интересующему пользователя понятию, просмотра и фильтрации списков выбранных ИО, навигации по конкретным ИО, а также просмотра описания выбранного информационного ресурса.

Список ИО отображается в виде страницы, содержащей набор ссылок на эти объекты. Для больших списков формируется составная страница, включающая список страниц с элементами навигации по этому списку.

Вся информация о конкретном объекте и его связях отображается в виде HTML-страницы (рис. 4), формат и наполнение которой зависят от свойств понятия, экземпляром которого является данный объект, и заданного для него шаблона визуализации. При этом объекты, связанные с данным объектом, представляются на его странице в виде гиперссылок, по которым можно перейти к их детальному описанию.

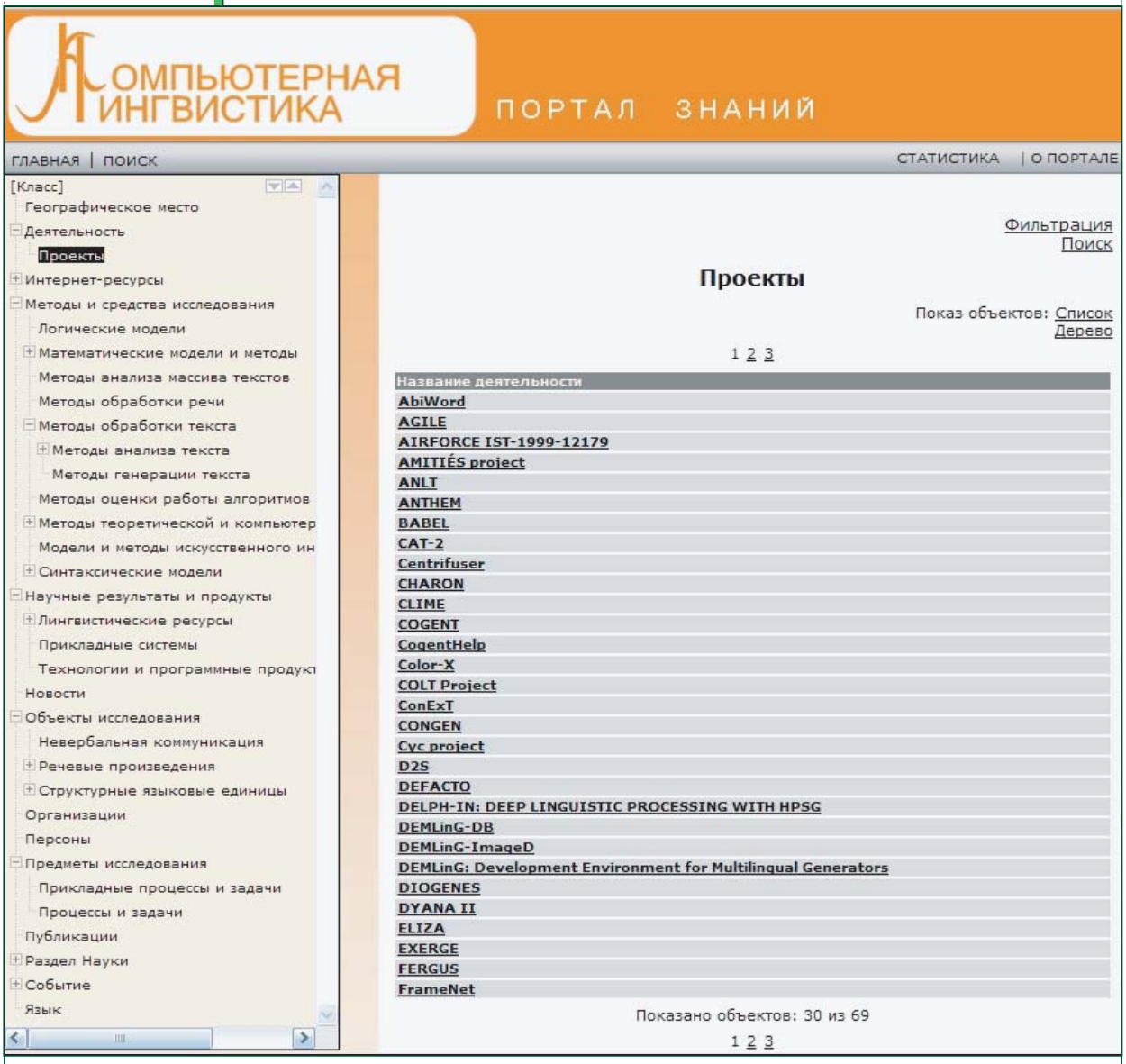
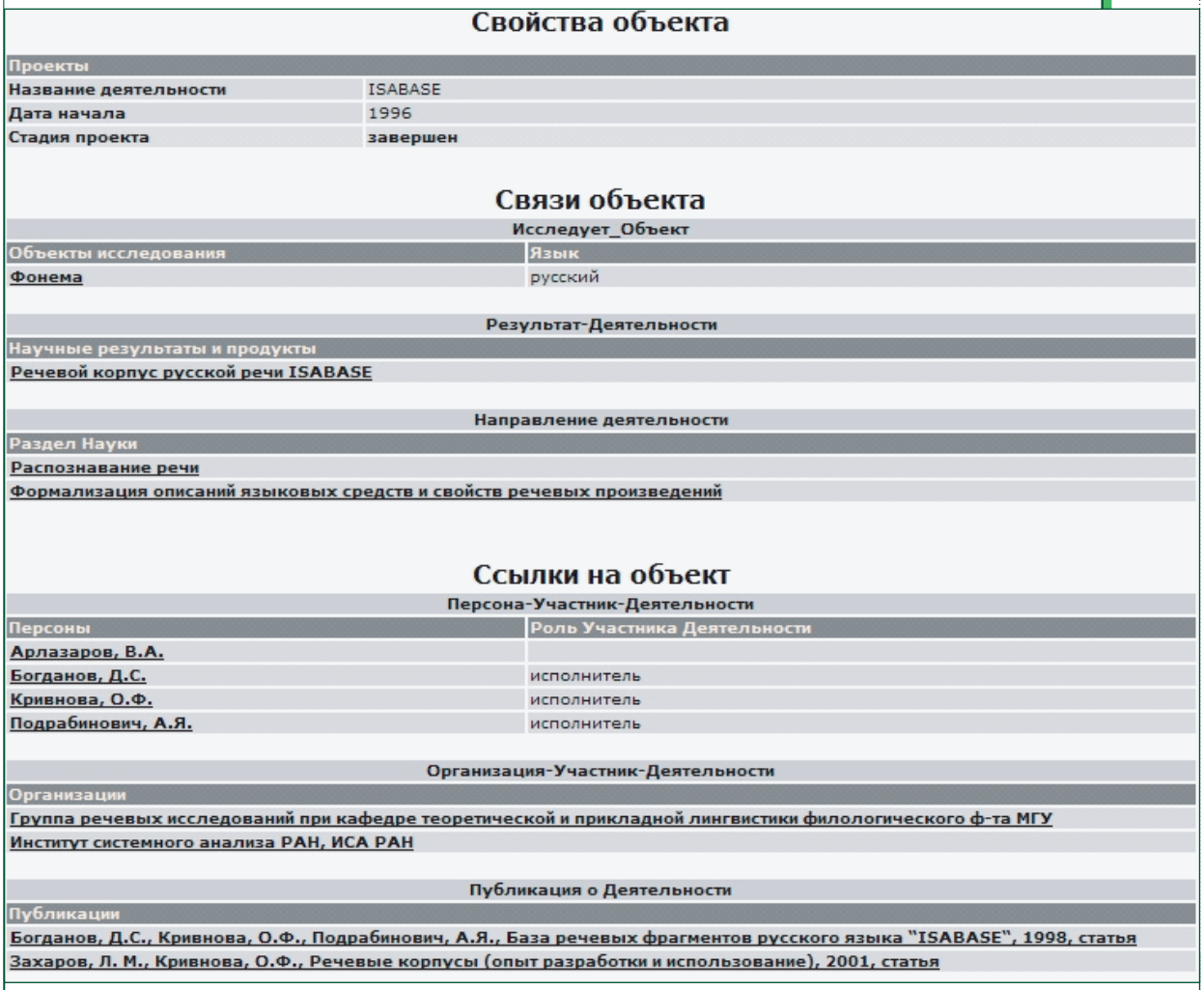


Рис. 3. Навигация по порталу знаний

Таким образом, навигация по данным портала представляет собой процесс перехода от одних информационных объектов к другим по заданным между ними связям.

Например, при просмотре информации о конкретном проекте (см. рис. 4) мы можем видеть значения его атрибутов и его связи с другими объектами. Используя представленные связи в качестве элементов навигации, можно перейти к просмотру подробной информации о научных результатах, полученных в ходе выполнения проекта, об участниках проекта, публикациях о нём и т.п.



Свойства объекта

Проекты	
Название деятельности	ISABASE
Дата начала	1996
Стадия проекта	завершен

Связи объекта

Исследует_Объект

Объекты исследования	Язык
Фонема	русский

Результат-Деятельности

Научные результаты и продукты
Речевой корпус русской речи ISABASE

Направление деятельности

Раздел Науки
Распознавание речи
Формализация описаний языковых средств и свойств речевых произведений

Ссылки на объект

Персона-Участник-Деятельности

Персоны	Роль Участника Деятельности
Арлазаров, В.А.	
Богданов, Д.С.	исполнитель
Кривнова, О.Ф.	исполнитель
Подрабинович, А.Я.	исполнитель

Организация-Участник-Деятельности

Организации
Группа речевых исследований при кафедре теоретической и прикладной лингвистики филологического ф-та МГУ
Институт системного анализа РАН, ИСА РАН

Публикация о Деятельности

Публикации
Богданов, Д.С., Кривнова, О.Ф., Подрабинович, А.Я., База речевых фрагментов русского языка "ISABASE", 1998, статья
Захаров, Л. М., Кривнова, О.Ф., Речевые корпусы (опыт разработки и использование), 2001, статья

Рис. 4. Представление информационного объекта и его связей

При переходе по конкретной связи любого информационного объекта мы можем получить достаточно большой список объектов (например, список людей, работающих в некоторой организации). В связи с этим был введён механизм фильтрации списков информационных объектов, который позволяет, например, отфильтровать множество публикаций как по дате публикации, так и по описываемому научному результату или объекту исследования.

3.2. Поиск в терминах предметной области портала

При поиске информации пользователю предоставляется возможность задания запроса в терминах предметной области портала. При этом пользователь должен выбрать понятие, к которому относятся искомые информационные объекты, и определить ограничения, которым должны удовлетворять атрибуты выбранного понятия и его связи с другими понятиями.

Ограничения на отдельные атрибуты интерпретируются как конъюнкция условий. Допустимые ограничения для атрибута зависят от типа его значений. Так, например, для атрибутов

типа «integer» и «date» задаётся точное значение или допустимый интервал значений.

Пользователю также предоставляется возможность задать условия на значения атрибутов объектов, связанных с искомым объектом. При этом могут быть заданы ограничения и на значения атрибутов соответствующих отношений.

Например, для получения ответа на вопрос «Найти проекты, выполнявшиеся после 1995 года, в которых принимал участие В.Л. Арлазаров и исследовались русские фонемы» пользователь должен выбрать в дереве онтологии понятие «Проект», а затем в автоматически сгенерированной поисковой форме задать ограничения на значения соответствующих атрибутов объектов и отношений. В результате будет сформирован следующий запрос:
Понятие *«Проект»*
Атрибут *«Дата начала»(>=1995)*
Отношение *«Исследует_Объект»*
Атрибут отношения *«Язык» = «русский»*
Понятие *«Объект исследования»*
Атрибут *«Название» = «Фонема»*
Отношение *«Персона-Участник-Деятельности»:*
Понятие *«Персона»*
Атрибут *«Фамилия» = «Арлазаров»*
Атрибут *«Инициалы» = «В.Л.»*

Заключение

В статье описан специализированный интернет-портал, обеспечивающий содержательный доступ к знаниям и информационным ресурсам по компьютерной лингвистике.

Портал представляет знания об основных разделах компьютерной лингвистики, о её предмете и объектах исследования, используемых в ней моделях и методах, разработанных системах, алгоритмах и лингвистических ресурсах, содержит информацию об учёных, сообществах, организациях, вовлечённых в процесс исследований по компьютерной лингвистике, о выполняемых проектах в этой области. Пользователи портала имеют доступ не только к текстовым ресурсам по КЛ, но и к ресурсам, представляющим реальные прикладные системы, технологии и программные продукты для обработки естественного языка, словари и лингвистические базы данных.

Благодаря тому, что систематизация и структуризация знаний и данных по компьютерной лингвистике выполнена на основе онтологии, вся информация на портале представлена в виде сети взаимосвязанных информационных объектов. Доступ к знаниям и данным портала осуществляется путём навигации по дереву понятий онтологии и его информационному пространству, а также через средства содержательного поиска.

Портал знаний по компьютерной лингвистике функционирует и доступен по адресу http://uniserv.iis.nsk.su/cl/. В планах авторов — дальнейшее развитие онтологии компьютерной лингвистики, сбор и интеграция в контент портала новых лингвистических знаний и информационных ресурсов.

Литература

1. Загорулько Ю.А., Боровикова О.И., Загорулько Г.Б. Организация содержательного доступа к информационным ресурсам на основе онтологий // Электронные библиотеки: перспективные методы и технологии, электронные коллекции. Тр. 9-й Всероссийской научной конф. RCDL'2007. Переславль-Залесский: Изд-во «Университет города Переславля», 2007. Т.1. С. 217–224.

2. Guariano N., Giaretta P. Ontologies and Knowledge Bases. Towards a Terminological Clarification // Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing. Amsterdam: IOS Press, 1995. P. 25–32.

3. Загорулько Ю.А., Боровикова О.И. Технология построения онтологий для порталов знаний по гуманитарным наукам // Труды Всероссийской конференции с международным участием «Знания-Онтологии-Теории» (ЗОНТ-07). Новосибирск, 2007. Т.1. С. 191–200.

Загорулько Юрий Алексеевич,
кандидат технических наук, заведующий лабораторией
Института систем информатики им. А.П. Ершова СО РАН.
e-mail: zagor@iis.nsk.su.

Соколова Елена Григорьевна,
кандидат филологических наук,
доцент Российского государственного гуманитарного университета, Москва.

Кононенко Ирина Семёновна,
научный сотрудник ИСИ СО РАН.

Загорулько Ирина Борисовна,
научный сотрудник ИСИ СО РАН.

Боровикова Олеся Игнатьевна,
младший научный сотрудник ИСИ СО РАН.

Вопросы речевых технологий на XVI Международном конгрессе фонетических наук (2007 г.)

Е. В. Шаульский

В обзоре излагается содержание сообщений, представленных на XVI Международном конгрессе фонетических наук (Саарбрюккен, 2007 г.) в секции речевых технологий.

6–10 августа 2007 г. в Саарбрюккене (ФРГ) состоялся XVI Международный конгресс фонетических наук, в котором приняли участие 535 фонетистов из 39 стран. В настоящем обзоре рассматриваются доклады секции речевых технологий, представленные в сборнике трудов и материалов конгресса [1].

В работе *С. Крстулович, А. Хунекке и М. Шрёдера* (Саарбрюккен) [2] обсуждаются результаты использования скрытых марковских моделей (Hidden Markov Modelling, HMM) при синтезе экспрессивной немецкой речи. Авторы рассматривают преимущества системы синтеза, основанной на HMM, перед другими системами синтеза речи — формантно-ориентированными (formant-based) и конкатенативными (unit-selection based): в отличие от первых, HMM-системы обеспечивают высокое качество синтезируемого голоса, а в отличие от вторых, не так сильно зависят от лежащей в основе голосовой базы данных. В случае с экспрессивной речью важнейшим свойством системы синтеза является способность синтезировать просодические особенности речи. Для решения этой задачи авторы, во-первых, использовали базу данных нейтрального немецкого голоса — BITS German speech synthesis corpus, и, во-вторых, создали небольшую базу данных экспрессивных высказываний, имитирующих речь немецкого футбольного комментатора (Bundesliga database). Эти высказывания были подвергнуты параметризации с использованием скрытых марковских моделей, и их просодические характеристики были «наложены» на нейтральный голос из первой базы данных. Результаты эксперимента — синтезированные с помощью данной системы экспрессивные немецкие предложения — авторы оценивают как «в целом приемлемые», хотя и отмечают ряд недостатков их просодического оформления: несмотря на сохранение оригинального уровня тона, синтезированный экспрессивный голос звучит «сдавленно» и «скрипуче», темп синтезированной речи заметно ниже, чем в оригинале, а девиация

основной частоты (F0 deviation) у синтезированного голоса составляет всего 33 % от соответствующего параметра оригинала. В электронной версии Материалов конгресса к данной статье приложены аудиофайлы оригинальной и синтезированной немецкой речи, и читатель может самостоятельно оценить, насколько успешно используемая авторами система синтеза справляется со своей задачей.

Просодическому моделированию немецких слов посвящён и доклад *У. Хиршфельд, Р. Хоффмана и Ф. Ланге* (ФРГ) [3], которые занимаются созданием произносительного словаря немецкого языка (Aussprachwörterbuch), содержащего звуковой модуль. Система синтеза речи в таком словаре должна порождать произношение слов (заголовков словарных статей) на основании имеющейся фонетической транскрипции и приписанного каждому слову признака принадлежности к той или иной акцентной модели (включающей данные о соотношении длительностей гласных в слове, мелодическом контуре и т. п.). В докладе описан процесс совершенствования набора таких акцентных шаблонов для повышения качества синтезируемых слов.

М. Михкла (Таллинн) [4] указывает на значимость морфологических и синтаксических факторов в определении длительности сегментов при синтезе речи на эстонском языке. Автор произвёл моделирование длительности сегментов речи дикторов эстонского радио при помощи статистических методов (линейной регрессии и нейронных сетей), введя в исходные данные информацию о частеречной принадлежности слова, его синтаксической роли и морфологических признаках. Результатом этого стало уменьшение числа ошибок при предсказании длительности сегментов, что доказывает необходимость учёта не только фонетических, но и грамматических факторов при синтезе речи на языке с богатой морфологией (каким является эстонский).

К. Барткова и Д. Жуве (Ланьон, Франция) [5] рассматривают проблемы обнаружения иностранного акцента при автоматическом распознавании речи. Известно, что распознавание речи с иностранным акцентом является одной из наиболее сложных задач автоматического распознавания. Использование моделей, ориентированных на родной язык, не может достаточно хорошо справиться с речью иностранца; с другой стороны, модели, построенные на материале не только родного, но и иностранных языков, показывают худший результат в распознавании речи носителей языка. Авторы работы предлагают предварительно автоматически определить степень иностранного акцента, чтобы затем, с учётом полученных данных, применять ту или иную модель распознавания. Была создана база данных из французских слов, произнесённых носителями французского языка, а также носителями английского, немецкого и испанского языков. Для автоматического определения акцента производилось три цикла декодирования: в каждом случае использовалась контекстно-зависимая модель распознавания для французского языка, а также для одного из трёх других языков — немецкого, английского и испанского, после чего вычислялось соотношение сегментов, распознанных как французские, к общему их числу; в зависимости от величины этого коэффициента определялась степень иностранного акцента. Для распознавания «сильно акцентированной» речи в дальнейшем используется специальная модель, адаптированная для иностранного языка (foreign-adapted model). Если же степень акцента не очень высока, распознавание производится при помощи модели, ориентированной на родной язык (native model).

На более масштабном материале исследуют иностранные акценты во французском языке *Б. Виеру-Думулеску и её соавторы* (Орсе, Франция) [6]. Ими был создан корпус французских текстов, прочитанных носителями французского, арабского, английского, немецкого, итальянского, португальского и испанского языков (по 6 человек от каждого

языка). После этого было произведено измерение некоторых сегментных признаков, как то: частоты формант гласных, длительность согласных и степень их звонкости, а также наличие или отсутствие факультативного [ə] в финальной позиции. Затем для каждого слова были определены произносительные варианты с учётом фонетических особенностей того или иного акцента, например, оглушения звонких, фрикативизации смычных, различных реализаций /r/, назальных гласных и т. п., и создан своего рода словарь произносительных вариантов. В результате выявились произносительные «предпочтения» носителей того или иного языка, говорящих по-французски (арабы не различают /e/ и /i/, немцы и англичане произносят глухие согласные с придыханием, испанцы не делают различия между /b/ и /v/, и т. п.) и определена их частотность.

Ф. Була де Мареуиль, М. Адда-Деккер и С. Вёрлинг (Орсе, Франция) [7] исследовали реализацию ртовых и носовых гласных в северной и южной разновидностях французского языка. Для этого они создали корпус данных из записей речи 12-ти географических пунктов северной и южной Франции. Обследование этого корпуса (с использованием методов автоматической обработки речи) позволило получить количественные данные о реализации гласных фонем, подтвердившие давние наблюдения французских диалектологов и социолингвистов: более переднее произношение /ё/ (вплоть до [œ]) характерно для северной Франции, тогда как на юге более частотно расщепление носовых гласных на (носовой или ртовый) гласный + носовой согласный.

Доклад *И. Лапри и А. Бонно* (Франция) [8] посвящён построению стимулов перцепции при помощи «синтеза с копированием» (copy synthesis). Для экспериментов по восприятию требуются речевые стимулы с изменяемым акустическим содержанием. Одну из возможностей для создания таких стимулов предоставляет система формантного синтеза, позволяющая вручную редактировать параметры синтезируемого звука. Авторами доклада была предложена система «синтеза с копированием», развивающая возможности формантного синтезатора. Синтез с копированием включает два этапа. Вначале производится вычисление основной частоты и определение параметров источника, в частности, соотношение голоса и шума. Второй шаг — задание формантных амплитуд. В данной работе намечаются пути развития системы синтеза с копированием в двух направлениях: автоматическое отслеживание динамики частоты формант (automatic formant tracking), при котором учитывается взаимозависимость формантных кривых, а также задание формантных уровней с использованием алгоритма тонального маркирования.

Проект системы синтеза речи по артикуляционным данным «Ouisper» представлен в работе *Т. Уэбера, Ж. Шолле, Б. Денби, М. Стоун и Л. Зуари* (Париж — Балтимор) [9]. Синтез речи в системе «Ouisper» должен осуществляться на основании артикуляционных данных, полученных из ультразвуковых изображений речевого тракта и видеозаписей движения губ говорящего. Применение HMM-моделирования к корпусу аудиовизуальных данных (также использовался алгоритм Unit-Selection) позволяет соотнести акустические данные с артикуляционными и построить систему синтеза речи, которая может быть использована в качестве альтернативы трахео-пищеводной речи больных раком гортани, в ситуациях, требующих сохранения тишины, а также для голосового общения в обстановке повышенного шума. Обсуждаются проблемы предварительной обработки ультразвуковых изображений, извлечения релевант-

ной информации из положения языка и губ, автоматической сегментации акустического сигнала. Авторы утверждают, что на данном этапе система продуцирует фонетическую транскрипцию только на основании видеосигнала (т. е. без обращения к аудиоданным) с точностью 50 %. В дальнейшем для решения задачи синтеза предстоит разработать систему «виртуальной просодии», для чего предполагается использование автоматизированной модели извлечения «просодических шаблонов» из данных корпуса.

Группа исследователей из Германии, Италии, Франции и Израиля [10] представила доклад, посвящённый проблемам соотношения тона и длительности в эмоциональной и аффектированной речи. Ставится под сомнение традиционная точка зрения, что тон играет более важную роль в маркировании эмоциональных состояний, чем другие просодические признаки. Авторы доклада провели следующее исследование. Была создана речевая база данных высказываний детей при общении с собакой-роботом AIBO. Полученные высказывания затем классифицировались как нейтральные либо содержащие какую-либо эмоцию (злость, нежность, эмпатия). После этого при помощи системы ESPS было осуществлено автоматическое извлечение данных о движении частоты основного тона в этих высказываниях, после чего полученные данные были скорректированы вручную одним из авторов на основе принципа «сглаживания и адаптации к человеческому восприятию», для того чтобы исключить влияние модуляций фонации (скрипучий голос, ларингализация и т. п.) на тональный контур. Таким образом, были получены два корпуса данных: первый — из автоматически определённых значений тонального движения (*aut*), второй — из тех же значений, подвергшихся «ручной» коррекции (*corr*). Для каждого из этих корпусов было вычислено соотношение параметров тона (F0) и длительности (DUR) в определении принадлежности высказывания к той или иной эмоции. Оказалось, что параметр F0 для *aut* более существен, чем для *corr*, тогда как DUR играет более важную роль для *corr*, чем для *aut*. Вместе с тем это различие не является «отчётливо выраженным». Авторы не дают ответа на вопрос, свидетельствуют ли полученные ими результаты о меньшей значимости тонального фактора в эмоциональной речи или же лишь об ошибках в автоматическом определении высоты тона, однако отмечают, что их выводы подчёркивают важность параметра длительности в общем комплексе просодических признаков.

Э. Лазарчик (Саарбрюккен) [11] посвятила сообщение изучению положений гортани для задач синтеза речи в артикуляторной модели. Ею было исследовано влияние положения гортани на качество гласных: во-первых, при подъёме гортани повышаются формантные частоты гласного; во-вторых, от положения гортани зависит и качество голоса: более высокое положение гортани соответствует напряжённому голосу, более низкое — расслабленному. Далее изолированные гласные в естественном произношении, но произнесённые с разным положением гортани — нейтральным, повышенным и пониженным, — сравнивались с гласными, синтезированными при помощи трёхмерной артикуляционной модели вокального тракта, где положение гортани было соответствующим образом симулировано. Проведённые измерения частоты первых трёх формант естественных и синтезированных гласных показали, что изменение положения гортани влияет на частоты формант и в том, и в другом случае. Что касается качества голоса, то манипулирования высотой гортани в артикуляторной модели оказалось недостаточно, чтобы достичь характеристик, присущих естественной речи. Помимо этого, потребовалось изменение параметров возбуждения, которые в сочетании с положением гортани позволяют достичь искомого результата.

Ф. Алиас и М. П. Тривиньо (Барселона) [12] предлагают таблицу для оценки разборчивости речи на каталанском языке. Описана процедура построения сбалансированной таблицы с учётом частотности согласных фонем в каталанском языке. В результате сформирована

таблица из 40 четырёхсловных списков, из которых 20 построены на изменении последней согласной фонемы, а другие 20 — на изменении первой.

В докладе *Дж. Уэллса* (Лондон) [13] обсуждаются вопросы использования фонетических символов в различных компьютерных приложениях (текстовых редакторах, программах электронной почты, веб-страницах и пр.). За последние годы широко распространился единый формат кодирования символов — Юникод, что позволяет и для фонетических символов применять единые методы кодирования, отказываясь от разнообразных шрифтов, не соответствующих международному стандарту. Одной из задач остаётся облегчение клавиатурного ввода специальных символов, для чего имеется ряд способов: Alt+номер, Таблица символов, Alt+X, специальные раскладки клавиатуры.

Литература

1. ICPhS 2007 — Proceedings of 16th International Congress of Phonetic Sciences (6–10 August 2007, Saarbrücken, Germany) // Edited of Jürgen Trouvain and William J Barry. Saarbrücken, 2007. Электронная версия: http://www.icphs2007.de.

2. Krstulović S., Hunecke A., Schröder M. Investigating HMMs as a parametric model for expressive speech synthesis in German // ICPhS 2007. P. 2181–2184.

3. Hirschfeld U., Hoffmann R., Lange F. Prosodic modelling of synthesised German words // ICPhS 2007. P. 2205–2208.

4. Mihkla M. Morphological and syntactic factors in predicting segmental durations for Estonian text-to-speech synthesis // ICPhS 2007. P. 2209–2212.

5. Bartkova K., Jouvet D. Automatic detection of foreign accent for automatic speech recognition // ICPhS 2007. P. 2185–2188.

6. Vieru-Dumulescu B., Boula de Mareüil Ph., Adda-Decker M. Characterizing non-native French accents using automatic alignment // ICPhS 2007. P. 2217–2220.

7. Boula de Mareüil Ph., Adda-Decker M., Woehrling C. Analysis of oral and nasal vowel realisation in Northern and Southern French varieties // ICPhS 2007. P. 2221–2224.

8. Laprie Y., Bonneau A. Construction of perception stimuli with copy synthesis // ICPhS 2007. P. 2189–2192.

9. Hueber T., Chollet G., Denby B., Stone M., Zouari L. Ouisper: Corpus based synthesis driven by articulatory data // ICPhS 2007. P. 2193–2196.

10. Batliner A., Steidl S., Schuller B., Seppi D., Vogt T., Devillers L., Vidrascu L., Amir N., Kessous L., Aharonson V. The impact of F0 extraction errors on the classification of prominence and emotion // ICPhS 2007. P. 2201–2204.

11. Lasarcyk E. Investigating larynx height with an articulatory synthesizer // ICPhS 2007. P. 2213–2216.

12. Alías F., Triviño M. P. A phonetically balanced modified rhyme test for evaluating Catalan speech intelligibility // ICPhS 2007. P. 2197–2200.

13. Wells J. An update on phonetic symbols in Unicode // ICPhS 2007. P. 2225–2228.

Е.В. Шаульский
Аспирант филологического факультета МГУ им. М. В. Ломоносова.

Анкета на тему: нужна ли специализация «Речевые технологии» в российском вузе?

О.Ф. Кривнова

Появление компьютеров и их проникновение в разнообразные сферы социальной жизни привели к созданию и развитию особых направлений в компьютерных технологиях, которые связаны со звуковой речью. Нет необходимости специально доказывать, что устная речь представляет собой наиболее удобный и естественный способ общения человека с компьютером, не требующий специального обучения. Речевые технологии, получившие мощный импульс к развитию в 70–80е годы прошедшего века, сейчас уверенно завоёвывают новые позиции и в научном плане, и в различных практических сферах жизни как целого общества, так и отдельного человека.
Чтобы продемонстрировать это, достаточно привести перечень основных областей применения компьютерных продуктов, разработанных и разрабатываемых в сфере речевых технологий.

- Человекомашинные интерфейсы с устным вводом/выводом информации.
- Речевое управление компьютером и другими техническими устройствами (особенно в экстремальных, опасных для человека условиях).
- Информационно-справочные службы, позволяющие получать и выдавать различную информацию из базы данных в условиях, когда вопрос задаётся голосом (на транспорте, в области туризма, навигаторы по незнакомой местности, в медицине, банковской службе, в навигаторах сети Интернет).
- Эффективное кодирование, сжатие и распознавание речи в телекоммуникационных каналах передачи информации; сотовая связь, поисковые системы Интернет.
- Многоязычный устный ввод/вывод речевой информации с автоматическим переводом.
- Приспособления и компьютерные программы для помощи инвалидам (слепым, глухим, немым, парализованным людям).
- «Автоматическая машинистка» — диктовальная машина, которая распознаёт речевое сообщение и записывает его в обычном текстовом виде.

- Озвучивание корректур и исправление ошибок.
- Помощь в обучении иностранному языку (автоматические фонетические тренажёры, электронные словари со звуковой поддержкой).
- В лингвокриминалистике, включая борьбу с международным терроризмом (обеспечение защиты от несанкционированного доступа).
- В медицинской диагностике.
- В научных исследованиях: в компьютерных моделях искусственного интеллекта и фонетических механизмов звучащей речи; в описательной и экспериментальной фонетике.

В настоящее время речевые компьютерные продукты и приложения создаются для всех более-менее распространённых, или, как их называют, мировых языков: таких, как английский, немецкий, французский, испанский, итальянский, греческий, японский, китайский (в том числе создаются и промышленные компьютерные продукты, по крайней мере, для первых двух языков из этого списка). Специалисты, работающие в сфере речевых технологий, регулярно встречаются и обмениваются опытом, обсуждают проблемные вопросы на международных конференциях, наиболее представительными из которых являются ICSLP — Int. Conf. on Spoken Language Processing; ICASSP — Int. Conf. in Acoustics, Speech and Signal Processing; EUROSPEECH; INTERSPEECH, SPECOM — Межд. конференция «Speech and Computer». В последнее время заметна активизация речевых разработок в странах Восточной Европы, особенно в Чехии и Польше.

Русский язык, к сожалению, не входит пока что в разряд языков технологического будущего. Промышленных разработок нет ни в области синтеза речи, ни в области распознавания. Причин много, и они разные. Но, думается, что одна из них, может быть, самая важная, — это отсутствие продуманной концепции подготовки профессиональных кадров в области речевых технологий и, в частности, отсутствие такой специализации в вузах. Между тем во многих зарубежных университетах, институтах и научно-исследовательских центрах знания и навыки профессиональной работы в сфере речевых технологий можно получить, обучаясь на специализированных отделениях — в рамках Computer Science, Cognitive Science, Speech and Language Engineering, Speech Processing, Electrical and Computer Engineering и т.п.

Подготовка специалистов в области речевых технологий осложняется тем, что эта научная и прикладная область носит междисциплинарный характер. При разработке прикладных систем, работающих с устной речью, возникают сложные и разноплановые проблемы. В их решении участвуют учёные и специалисты из разных областей науки: лингвисты, физиологи, психологи, математики, физики, инженеры, специалисты в области компьютерной науки. Уже сейчас в отдельных российских вузах читаются лекционные курсы, ведутся практические занятия и научные семинары, имеющие отношение к проблематике речевых технологий. Однако кажется, что наступило время, когда необходимо переходить от усилий отдельных учебных подразделений и учёных энтузиастов к разработке действенной кадровой политики в этой важной социальной сфере образования. Раньше отставание в этой области можно было «списать» на отсутствие необходимой компьютерной техники, сейчас — на финансовый кризис, но, наверное, есть и другие глубинные причины.

Редколлегия журнала «Речевые технологии» приглашает специалистов в области речевых технологий и других коллег, занимающихся и интересующихся проблемами устной речи, принять участие в небольшом анкетном опросе, посвящённом задачам подготовки специалистов по речевым технологиям.

Мы сознательно ограничились наиболее общими и принципиальными, на наш взгляд, вопросами в надежде на конструктивный диалог, в том числе и на расширение анкеты дополнительными вопросами, которые остались за пределами её первого варианта.

Вопросы анкеты приводятся ниже. После ответов наших респондентов и в завершение дискуссии будут подведены итоги обсуждения и проанализированы перспективы и направления усилий, которые помогли бы улучшить ситуацию с развитием речевых технологий, в особенности на материале русского языка.

ВОПРОСЫ АНКЕТЫ

(Отвечая на вопросы, напишите, пожалуйста, сначала Ваши данные: ФИО, место работы, должность и научную степень, если она есть).

1. Считаете ли Вы необходимым или целесообразным введение специализации «Речевые технологии» в перечень специализаций вузовского образования в России?

2. На базе каких вузов/факультетов/отделений целесообразно развивать такую специализацию? С какими профильными дисциплинами и в каком объёме?

3. Преподаете ли Вы сами что-либо, имеющее отношение к речевым технологиям? В каком вузе, на каком факультете и курсе? В каком объёме учебных часов?

4. Если Вы преподаёте что-либо, имеющее отношение к речевым технологиям, то какие курсы Вы читаете, есть ли утвержденная программа этих курсов, пишутся ли по их тематике под Вашим руководством курсовые, дипломные или диссертационные работы?

5. Знаете ли Вы других специалистов, которые преподают что-либо, имеющее отношение к речевым технологиям? Знакомы ли Вы с их учебными программами? Считаете ли Вы целесообразным/полезным согласование программ по речевым курсам, которые читаются в разных российских вузах разными специалистами?

6. Приглашаете ли Вы других специалистов по речевым технологиям читать какие-то отдельные разделы Вашего курса? Считаете ли Вы полезным/целесообразным приглашение/обмен преподавателями или учебными курсами в сфере речевых технологий между разными российскими вузами?

7. Принимали/принимаете ли Вы участие в каких-либо проектах, имеющих отношение к речевым технологиям? Если да, то укажите, пожалуйста, в каких именно, в какие сроки, на материале каких языков. Используете ли Вы этот практический опыт в своей преподавательской деятельности?

8. Имеете ли Вы контакты с какими-либо разработчиками современных речевых технологий и конкретных приложений? Пользуетесь ли Вы их помощью в своей преподавательской работе? Если да, то в какой форме? Считаете ли Вы полезной, целе-

сообразной организацию сотрудничества между учебными и научно-исследовательскими речевыми центрами, в том числе коммерческой направленности, в форме организации студенческой практики, стажёрских мест для потенциальных молодых специалистов по речевым технологиям?

9. Знакомы ли Вы с тем, как ведётся подготовка специалистов по речевым технологиям за рубежом? Какой из зарубежных учебных и научных центров кажется Вам наиболее перспективным и продвинутым в подготовке квалифицированных многопрофильных разработчиков-речевиков?

10. Считаете ли Вы полезным и целесообразным (несмотря на возможные финансовые трудности) приглашение ведущих зарубежных специалистов для чтения небольших (до полугода) курсов по основным направлениям в сфере речевых технологий?

11. Считаете ли Вы полезным и целесообразным (несмотря на возможные финансовые трудности) направлять в ведущие зарубежные научно-исследовательские речевые центры наиболее способных молодых специалистов из России для дополнительного обучения/приобретения конкретного опыта работы?

12. Считаете ли Вы полезным и целесообразным (несмотря на возможные финансовые трудности) приглашать ведущих зарубежных специалистов-речевиков для реализации конкретного речевого проекта (с нуля и под ключ, с набором команды из молодых российских специалистов) на достаточно длительный срок (не меньше года на контрактной основе)?

/Анкету подготовила
***О.Ф. Кривнова**,*
доктор филологических наук,
старший научный сотрудник
кафедры теоретической
и прикладной лингвистики МГУ/

Распознавание ключевых слов в потоке речи при помощи фонетического стенографа

В.В. Пилипенко

В статье рассматривается использование фонетического стенографа для распознавания ключевых слов в потоке речи. Для моделирования фонем используются Скрытые Марковские Модели. Ключевое слово задаётся последовательностью фонем в виде транскрипции слова. Приведены результаты поиска ключевых слов в потоке речи большого количества дикторов. Предложенный подход может использоваться для поиска речевой информации в огромных массивах данных.

Введение

В связи со всё более активным использованием естественного интерфейса (и, в частности, голоса) для общения с техникой возросло и значение аудиозаписи как носителя информации. Появилась потребность в системах, способных быстро и эффективно обслуживать аудиоархивы и находить нужную информацию в большом объёме записи. Для этой цели предложено использовать алгоритмы поиска ключевых слов в потоке речи.

Задачей поиска ключевых слов является нахождение заданных фрагментов (это могут быть отдельные слова или целые фразы) в потоке речи. Первоначально для задания фрагментов использовались отрезки произнесённой речи, при этом по нескольким произнесениям формировался эталон ключевого слова. Неудобство такого метода проявлялось в том, что для введения в систему нового ключевого слова необходимо заранее его произнести или вырезать из известного потока речи.

Современные алгоритмы поиска ключевых слов используют задание ключевых слов последовательностью фонем или других элементарных единиц. При этом может использоваться преобразователь графема-фонема в соответствии с правилами данного языка, и тогда ключевое слово задаётся текстом слова или фразы, что значительно расширяет область применения такой системы.

Широкое применение получили алгоритмы, в которых для моделирования элементарных единиц уровня фонемы применяются Скрытые Марковские Модели (СММ).

Для поиска ключевых слов используются те же подходы, что и для распознавания слитной речи.

Модификация касается способа задания слов, отсутствующих в словаре системы. Предложено два способа задания неизвестных слов:
1) моделирование незнакомых слов произвольными последовательностями фонем;
2) использование Гауссовской Смеси Моделей (*Gaussian Mixture Model GMM)* для моделирования фонового потока речи.

В данной статье рассматривается первый способ задания незнакомых слов. Для этого используется концепция фонетического стенографа [1], [2].

1. Базовая система распознавания слитной речи

В данной работе используется инструментарий HTK [3] на основе CMM. При помощи инструментария HTK построены акустические и лингвистические модели системы. Для распознавания речи был разработан программный комплекс, совместимый с акустическими и лингвистическими моделями HTK.

1.1. Предварительная обработка речевого сигнала

Речевой сигнал преобразуется в последовательность векторов признаков с интервалом анализа 25 мс и шагом анализа 10 мс. Вначале речевой сигнал фильтруется фильтром высоких частот с характеристикой $P(z)=1-0.97z^{-1}$. Затем применяется окно Хэмминга и вычисляется быстрое преобразование Фурье. Спектральные коэффициенты усредняются с использованием 26 треугольных окон, расположенных в мел-шкале, и вычисляются 12 кепстральных коэффициентов.

Логарифм энергии добавляется в качестве 13-го коэффициента. Эти 13 коэффициентов расширяются до 39-мерного вектора параметров путём дописывания первой и второй разностей от коэффициентов, соседних по времени. Для учёта влияния канала применяется вычитание среднего кепстра.

1.2. Акустическая модель

В качестве акустических моделей используются CMM. 56 украинских контекстно-независимых фонем моделируются тремя состояниями Марковской цепи без пропуска. Используется диагональный вид Гауссовских функций плотности вероятности.

Редко встречающиеся фонемы моделируются 64 смесями Гауссовских функций плотности вероятности, более часто встречающиеся фонемы моделируются большим числом смесей, наиболее часто встречающиеся фонемы используют 1024 смесей.

Словарь транскрипций создаётся автоматически из орфографического словаря с использованием контекстно-независимых правил.

2. Акустическое и текстовое наполнение

2.1. Обучающая выборка

Обучение производилось на выступлениях депутатов Верховной рады Украины, записанных через телевизионную сеть. Парламентская речь характеризуется некоторыми особенностями.

Это спонтанная речь. Встречаются отдельные доклады, зачитываемые по подготовленному заранее тексту, однако мало дикторов в точности придерживается этого текста.

Из-за ограничения во времени выступления многих дикторов произносятся в слишком быстром темпе.

Часто речь эмоционально окрашена.

В основном, записи состоят из непрерывных выступлений дикторов, но в них встречаются реплики ведущего заседания или других депутатов.

Качество записи достаточно высокое, поскольку каждое депутатское место оснащено микрофоном.

Для обучения использовались записи длиной в 250 тыс. сек., в которых встретилось около 495 тыс. слов. Всего было записано 208 дикторов.

Обучение производилось на предварительно размеченной выборке. Для этого запись выступления автоматически разбивалась на фразы из нескольких слов, ограниченные паузами больше 400 мс. Среднее количество слов в одной фразе оказалось равным пяти.

Каждой фразе оператором ставилась в соответствие метка в виде текста из стенограммы. Затем автоматически производилось преобразование текста в последовательность фонем в соответствии с контекстно-независимыми правилами украинского языка. Размеченная таким образом выборка использовалась для построения акустической модели.

2.2. Контрольная выборка

Распознавание производилось на выступлениях депутатов, записанных в отличные от обучающей выборки дни. Для распознавания использовались записи длиной в 60 тыс. сек., в которых встретилось 80 тыс. слов. Всего использовались записи 118 дикторов. Записи 36 дикторов не встретились в обучающей выборке. Таким образом, эти дикторы оказались неизвестными для системы распознавания.

2.3. Текстовый материал

Словарь был составлен из текстов стенограмм заседаний Верховной рады Украины. С официального сайта Верховной рады были загружены все стенограммы заседаний, начиная с 1991 года, что составило больше 100 МБ текста. Текст был модифицирован, для того чтобы убрать служебную информацию из стенограмм (например, аплодисменты), записать числа в текстовом виде, а также отделить русский текст от украинского.

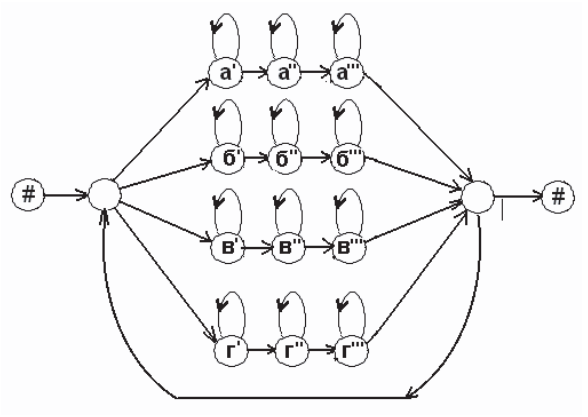


Рис. 1. Граф для произвольной последовательности фонем

3. Фонетический стенограф

Алгоритм фонетического стенографа позволяет строить последовательность фонем для речевого сигнала без использования какого-либо словаря. Для этой цели строится некоторая генеративная грамматика, которая может синтезировать все возможные модельные сигналы непрерывной речи для любой последовательности фонем. В рамках построенной модели строится алгоритм пофонемного распознавания для неизвестного сигнала. Используются те же контекстно-независимые модели фонем, как и в базовом распознавателе.

Надёжность найти фонему на правильном месте для известной реализации равна приблизительно 85%.

4. Результаты экспериментов по распознаванию ключевых слов в потоке слитной речи

Эксперименты проводились на описанной контрольной выборке.

Ключевые слова описывались последовательностью фонем заданной длины от 2 до 12 фонем. Для данной длины из словаря выбиралось 30 ключевых слов. К сожалению, для длин 2, 11 и 12 в тестовом корпусе не удалось выбрать достаточное количество записей, и в данном случае было выбрано около 20 ключевых слов. Всего было отобрано 309 ключевых слов.

Для каждого ключевого слова из тестового корпуса выбиралось от 15 до 100 записей фраз, в которые это ключевое слово обязательно входило. На данном материале подсчитывался процент *ложного отказа* (False Rejection) как доля случаев, когда ключевое слово не было распознано.

Кроме того, выбиралась выборка длиной в 1000 слов, в которую ключевое слово гарантированно не входило. На данном материале подсчитывался процент *ложного срабатывания* (False Alarm) как доля случаев, когда происходило срабатывание алгоритма распознавания ключевого слова.

Алгоритм содержит коэффициент, позволяющий регулировать соотношение между процентами *ложного отказа* и *ложного срабатывания.* Оптимальный коэффициент был выбран из условия минимума суммы этих процентов. При необходимости можно выбрать другое значение коэффициента, отдавая предпочтение тому или иному сценарию использования системы.

Таблица 1

Надёжность распознавания ключевых слов

Число фонем в ключевом слове	Процент ложного отказа	Процент ложного срабатывания
2	6.95	13.27
3	5.22	7.30
4	3.26	4.76
5	4.06	2.34
6	3.32	1.87
7	2.21	1.12
8	1.52	1.48
9	2.09	0.74
10	3.79	0.55
11	4.47	0.38
12	5.73	0.22
Вместе	3.67	3.02

В таблице 1 приведены результаты распознавания ключевых слов в зависимости от количества фонем в ключевом слове.

Оптимальное значение коэффициента зависит от длины слова, для более длинных слов его можно увеличить для получения лучших результатов.

Заключение

Статья описывает экспериментальную систему распознавания ключевых слов в потоке речи на основе фонетического стенографа. Проведены эксперименты по распознаванию. Коэффициент *ложного отказа* равен 3.67% при *ложном срабатывании,* равном 3.02%. Это позволяет надеяться, что данный алгоритм можно использовать в практических системах.

В дальнейшем предполагается рассмотреть комбинацию фонетического стенографа и модели фоновых слов в виде Гауссовской Смеси Моделей (*Gaussian Mixture Model GMM).*

Литература

1. Taras K. Vintsiuk. Generalized Automatic Phonetic Transcribing of Speech Signals // Труды пятой всеукраинской международной конференции «Оброблення сигналів і зображень та розпізнавання образів», Видання УАсОІРО, Київ, 2000, С.95–98.
2. Пилипенко В.В. Використання фонетичного стенографа при розпізнаванні мовлення з великих словників // Тезисы 12-й международной конференции «Автоматика-2005», Харьков, 2005, с.73.
3. Young S., Evermann G., Kershaw D., Moore G., Odell J., Ollason D., Valtchev V., Woodland P. The HTK Book. — Cambridge University Engineering Department, 2002.

В.В. Пилипенко,
сотрудник Международного научно-учебного центра информационных технологий и систем. г. Киев, Украина.
E-mail: valery_pylypenko@mail.ru.

Адаптивный алгоритм принятия решения «ТОН–НЕ ТОН», синхронный с основным тоном

И.А. Архипов,
кандидат технических наук

В.Б. Гитлин,
доктор технических наук

Д.А. Лузин

Признак «ТОН–НЕ ТОН» (Т/НТ) указывает на наличие или отсутствие вокализации в речевом сигнале. Он определяет способ образования звука [1] и служит одним из признаков параметрического описания речи. Его точная оценка необходима в системах анализа и синтеза речи [1], [2], [3].

Основными признаками, на основе знания которых принимается решение Т/НТ, служат следующие признаки [4].

1. *Энергия звука в различных областях спектра*: для вокализованных звуков она сосредоточена в низкочастотном диапазоне, для невокализованных — в высокочастотном. Энергия вокализованных звуков сконцентрирована в формантных областях, энергия невокализованных — распределена по спектру более равномерно [1], [2], [3].

2. *Энергия вокализованных звуков пульсирует с частотой основного тона* (ОТ), невокализованных — более равномерна, кроме взрывных /п/, /т/, /к/ и аффрикат /ц/, /ч/ [5], [6].

3. *Распределение вероятностей* мгновенных значений сигнала невокализованных звуков близко к гауссовскому закону, распределение для вокализованных звуков отлично от гауссовского. Отсчёты вокализированного сигнала существенно коррелированы между собой, корреляция отсчётов невокализированного сигнала слабее [4], [7].

4. *Частота пересечений нуля* сигналом вокализованных звуков ниже частоты пересечений нуля сигналом невокализованных звуков [1]. В общем слу-

чае частота пересечений нуля не служит надёжным признаком для принятия решения Т/НТ [4]. Это вызвано низкой помехоустойчивостью этого признака, широкой изменчивостью параметров фонового шума, большой зоной перекрытия распределений частоты переходов через нуль двух рассматриваемых классов («ТОН», «НЕ ТОН») [2].

Энергия вокализованных звуков выше энергии невокализованных звуков и пауз [1]. Алгоритмы принятия решения Т/НТ по энергии сигнала с фиксированным порогом имеют относительно низкую надёжность, поскольку принятие решения в существенной мере зависит от уровня сигнала и уровня шума [4]. Уровни сигнала и шума не остаются постоянными даже во время произнесения достаточно короткого текста [8]. Динамический диапазон акустического сигнала речи может достигать 80 дБ [1]. Для компенсации изменений сигнала по амплитуде используют адаптивный порог или нормализацию речевого сигнала [1].

Принятие решения по энергии в некоторой полосе частот, составляющей часть от полного спектра сигнала, позволяет учесть способ образования звука и тем самым повысить надёжность принятия решения [4], [9]. Однако ряд фрикативных и аспирированных шумных звуков, например, /ф/, /х/, имеют довольно мощные составляющие в низкочастотной части спектра, что может вызвать сбои систем принятия решения по энергии в полосе частот [4].

В работах [10], [11] Атал и Рабинер исследовали следующие признаки: нормированный коэффициент корреляции с единичной задержкой $R(1)$, первый коэффициент модели линейного предсказания a_1 при числе полюсов M=12 в ковариационном методе линейного предсказания и нормализованную ошибку линейного предсказания E_p.

Для вокализованной речи [11] $R(1)$ близко к единице, для невокализованной речи и шума $R(1)$ близко к нулю. Первый коэффициент линейного предсказания a_1 связан с $R(1)$ и зависит от порядка модели M, т.е. от формантной структуры звука. Нормализованная ошибка линейного предсказания E_p отражает степень близости спектра сигнала к спектру белого шума: чем спектр равномернее, тем ошибка больше. Для вокализованной речи E_p меньше, для невокализованной — больше.

Атал и Рабинер в работе [11] делают следующие выводы.

1. При принятии решения Т/НТ дополнительно появляются ошибки на интервалах паузы из-за изменчивости фонового шума, который различен для обучающей и контрольной выборок.
2. Большинство ошибок появляются на границе между классами. Ошибки возникают в случае, когда внутрь одной рамки анализа попадают два разных класса звуков.

Периодичность сигнала, связанную с основным тоном, можно оценить по виду спектра. Спектр вокализованных звуков неравномерен и концентрируется на гармониках. Спектр невокализованных звуков более равномерен [4]. Недостаток оценки периодичности сигнала по виду спектра — низкая помехоустойчивость, поскольку искажения и фоновые шумы могут существенно исказить истинный спектр [10].

Можно принимать решение Т/НТ по оценке периодичности сигнала путём перехода к анализу колебательности временной функции. Однако по данным работы [4] оценка степени колебательности временной функции речи не обеспечивает надёжного принятия решения Т/НТ.

В процессе выделения основного тона довольно часто вычисляют функции, которые могут служить мерой оценки периодичности, связанной с ОТ. Такими функциями могут быть:

значение максимума автокорреляционной функции, значение минимума разностной функции, величина пика кепстра и ряд других [12] [13], [14]. Недостатком данного способа принятия решения Т/НТ является зависимость указанных параметров от формантной структуры сигнала, от длины кадра анализа, от величины фонового шума и от ряда других факторов [7].

Повысить надёжность принятия решения Т/НТ можно путём увеличения количества признаков, по которым принимают решение. Повышение надёжности возможно в том случае, когда признаки независимы или, по крайней мере, слабо коррелированы относительно ошибок принятия решения Т/НТ [16]. Если решение Т/НТ принимают в многомерном пространстве признаков, то процедура принятия решения существенно усложняется, отсутствует наглядность представления распределений признаков, необходимо увеличение обучающей выборки. Для упрощения этой процедуры можно использовать методы теории распознавания образов [16]. Выбранная система признаков должна в совокупности обеспечить необходимую надёжность принятия решения при минимальной стоимости принятия решения.

Сегментацию речи на тональные интервалы выполняют синхронно [20], [21] и асинхронно с ОТ [1]…[3]. Асинхронная с ОТ обработка предполагает фиксированный размер кадра анализа. Согласно [17] оптимальная длительность интервала усреднения для энергии равна 10 мс. Текущий кадр анализа располагается случайным образом, и возможно попадание участков с разным типом возбуждения речевого тракта в один кадр. Решение о принадлежности данного кадра к какому-либо способу возбуждения во многом зависит от соотношения длительностей участков с разным способом возбуждения, попавших в данный кадр. На рис.1 показаны обобщённые схемы формирования признака Т/НТ синхронно и асинхронно с ОТ. На рис. 1*а* исходный сигнал сегментируют на тональные интервалы, а затем только тональные интервалы подвергают выделению ОТ. При сегментации речи асинхронно с ОТ кадры анализа имеют длительность, превышающую длительность периода ОТ, и следуют с перекрытием.

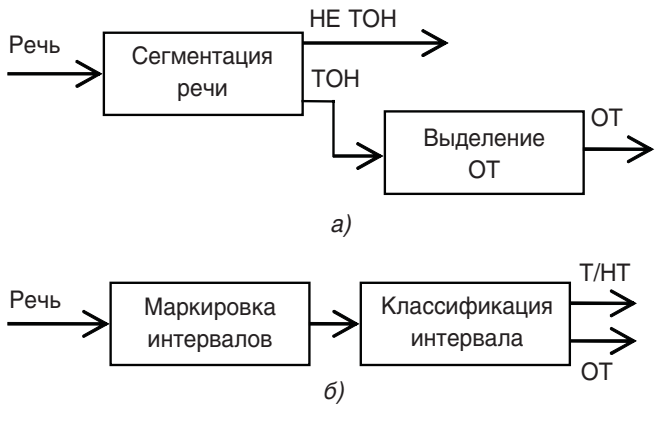


Рис.1. Способы классификации речи по признаку Т/НТ: а) асинхронно с ОТ; б) синхронно с ОТ

В обработке синхронной с ОТ кадры анализа привязаны к периодам ОТ. Привязка интервалов анализа к периодам ОТ позволяет избежать указанную выше неопределённость в расположении кадра анализа. Под кадром анализа здесь следует понимать участок сигнала между соседними марками, соответствующими началам новых периодов ОТ. Длительность каждого тонального интервала можно принимать за оценку периода ОТ. Кадры анализа следуют без перекрытия, за счёт чего существенно повышается скорость обработки.

Для простановки марок в началах периодов ОТ без предварительной сегментации на вокализированные и невокализированные интервалы необходимо использовать

локальный алгоритм выделения ОТ, в качестве которого выбран алгоритм, работающий по методу GS [18]. Синхронный с ОТ анализ ограничивает набор признаков, которые могут быть использованы для принятия решения Т/НТ, только такими, интервал вычисления которых может быть равен периоду ОТ (от 2 мс до 20 мс [1]). По этой причине из набора признаков, указанных выше, взяты только три признака [19]: нормированный коэффициент корреляции с единичной задержкой $R(1)$, логарифм частоты пересечения нулевого уровня и логарифм энергия сигнала в полосе частот 20…1500 Гц.

Нормированный коэффициент корреляции с единичной задержкой определяли следующим образом:

$$(\mathbf{R}\ 1) = \mathbf{K}_r \cdot \left(1 + \sum_{i=0}^{N-2} \mathbf{S}_i \cdot \mathbf{S}_{i+1} \Big/ \sum_{i=0}^{N-1} \mathbf{S}_i^{\ 2} \right), \qquad (1)$$

где K_r — нормирующий множитель, S_i — отсчёт входного речевого сигнала, не прошедшего этап предварительной обработки, N — число отсчётов на анализируемом периоде ОТ. Эксперименты показывают, что паузы в речи обычно заполнены слабыми, относительно случайными колебаниями, спектр которых зависит от спектра фонового шума. Поведение функции $R(1)$ в данном случае непредсказуемо.

На рис. 2*а*, 2*б* представлены осциллограмма слова «четыре» и функция $R(1)$ данного произнесения. На рис. 2*б* тональный и шумовой участки можно надёжно разделить по значениям функции $R(1)$. Поведение функции $R(1)$ на паузе (между марками 3–4) нестабильно и не позволяет классифицировать этот сегмент как невокализированный. Для лучшего разделения паузы и вокализованного сигнала по $R(1)$ необходимо приблизить спектр паузы к спектру невокализованных звуков. Для этой цели в работах [7], [11] предложено смешивать сигнал с шумом определённого уровня с подъёмом в сторону высоких частот.

Проведены эксперименты по оценке надёжности принятия решении Т/НТ по $R(1)$.

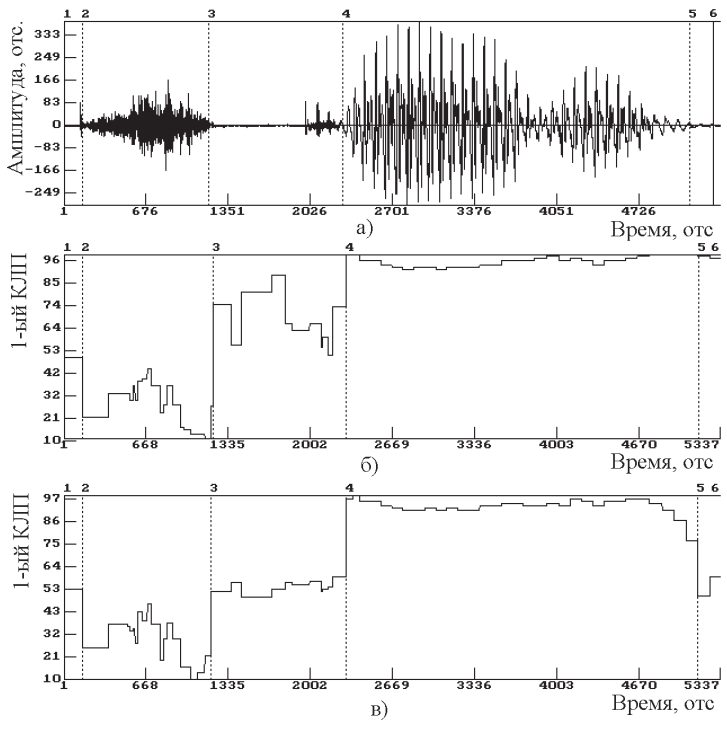


Рис. 2. Нормированный коэффициент корреляции с единичной задержкой:
а) осциллограмма слова «четыре»; б) функция нормированного коэффициента корреляции с единичной задержкой; в) функция нормированного коэффициента корреляции с единичной задержкой, вычисленного при добавлении шума с размахом 20 отсчётов

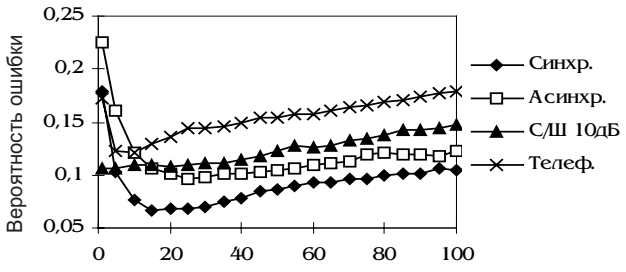


Рис. 3. Зависимость вероятности ошибки классификации Т/НТ по функции R(1) от уровня добавляемого шума

В качестве речевого материала использовали по одному произнесению фраз «Не видали мы такого невода», «Саша кусал сало», «На ухабе» и «Жирные сазаны ушли под палубу». В эксперименте принимали участие 12 дикторов (6 мужчин и 6 женщин).

Испытания проводили для чистого сигнала, для сигнала с аддитивным шумом при отношении С/Ш=10дБ и для сигнала, ограниченного полосой телефонного канала 300...3400Гц. Первоначально все фразы были вручную сегментированы на вокализированные и невокализированные сегменты.

К каждому произнесению был добавлен шум интенсивностью от 0 до 100 отсчётов уровней квантования с шагом через 5 отсчётов при максимуме сигнала в 32768 отсчётов. Нулевая интенсивность соответствует отсутствию шума. Результаты эксперимента показаны на рис. 3. Ошибку принятия решения Т/НТ определяли для синхронного с ОТ и асинхронного с ОТ методов принятия решения Т/НТ. Для телефонного сигнала и сигнала с аддитивным шумом при С/Ш=10 дБ анализ проводили синхронно с ОТ.

Кривая для синхронного способа вычисления признака $R(1)$ при всех уровнях добавляемого шума проходила ниже асинхронной кривой. Область минимума ошибки принятия решения Т/НТ была близка для всех типов исследованных сигналов, кроме сигнала с аддитивным шумом при С/Ш=10 дБ. В среднем, при добавлении оптимального значения шума синхронный с ОТ анализ по сравнению с асинхронным позволяет снизить вероятность суммарных ошибок классификации на 11%.

Энергия вокализованных звуков, как правило, выше энергии невокализованных звуков и пауз. Значение энергии определяли по формуле:

$$E = K_e \cdot \lg\left(\mathrm{e} + \sum_{i=1}^{N} x_i^{\,2}\right), \qquad (2)$$

где x_i — отсчёт речевого сигнала на выходе фильтра низких частот (ФНЧ) с частотой среза f_c, а K_e — нормирующий множитель.

На рис. 4. представлены осциллограммы произнесения слова «четыре», функция энергии исходного произнесения и функция энергии исходного произнесения, прошедшего через ФНЧ с частотой среза 1000 Гц. Энергию вычисляли синхронно с ОТ. Участок сигнала между марками 2–3 соответствует шумовому звуку «ч». Из рис. 4*б* видно, что звук «ч» имеет энергию, сравнимую с энергией вокализованных звуков. На рис. 4*в* энергия звука «ч» в значительной степени подавлена фильтром нижних частот. В данном случае можно легко отделить шипящий звук «ч» от вокализованных звуков. Эксперименты показывают, что с ростом частоты среза ФНЧ для значений, превышающих 1000 Гц, вероятность ошибки класси-

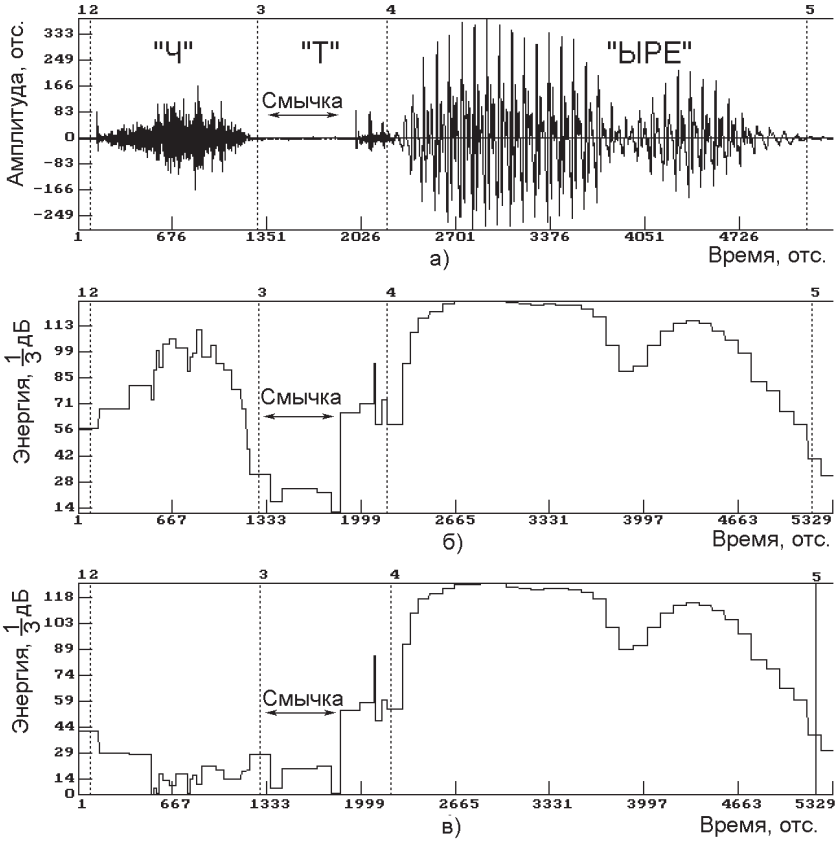


Рис. 4. Слово «четыре» (диктор — мужчина):
а) осциллограмма исходного произнесения; б) функция энергии исходного произнесения; в) функция энергии исходного произнесения, прошедшего ФНЧ с частотой среза 1000 Гц

фикации Т/НТ медленно монотонно возрастает. В последующих экспериментах мы ограничились частотой среза ФНЧ f_c=1500 Гц, определяемой требованиями алгоритма GS [22].

Вычисление энергии синхронно с ОТ приводит к снижению вероятности суммарной ошибки классификации по сравнению с асинхронным способом вычисления. Суммарная вероятность ошибки снижается на величину от 1,5% до 3,3% при минимальной ошибке классификации по энергии около 10% в зависимости от ширины полосы частот, в которой вычисляют энергию.

Частота пересечений нулевого уровня сигналом (ЧПН) имеет большой динамический разброс значений [1], [2], [23], вследствие чего предпочтительно в качестве признака классификации Т/НТ использовать логарифм частоты пересечения через ноль (ЛЧПН):

$$Z_{cr} = K_z \lg(M/T_0), \qquad (3)$$

где K_z — нормирующий коэффициент; M — количество пересечений нулевого уровня на периоде основного тона.

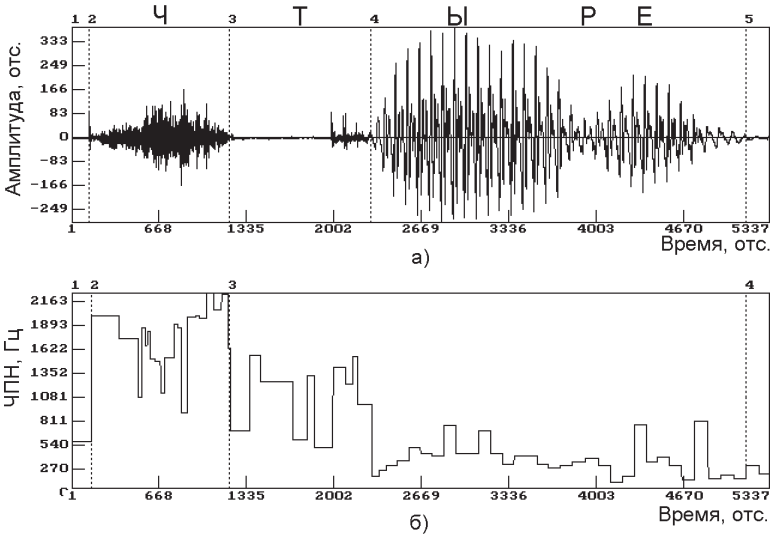


Рис. 5. Частота пересечения нулевого уровня речевого сигнала: а) осциллограмма слова «четыре» (диктор — мужчина); б) ЧПН сигнала

На рис. 5 изображены осциллограмма изолированного слова «четыре» и соответствующий ей график ЧПН. Марки 3, 4, 5 и 6 установлены на границах интервалов вокализации. Частота пересечений нуля вокализованных звуков ниже частоты пересечений нуля невокализованных звуков.

Из рис. 5*б* видно, что график признака ЧПН значительно изрезан, как на вокализированном, так и на невокализированном участках. Изрезанность графика ЧПН говорит о том, что короткие интервалы анализа при синхронном с ОТ способе вычисления ЧПН недостаточно сглаживают значения ЧПН, что приводит к указанному выше расширению динамического диапазона значений ЧПН. Распределения ЧПН вокализированных и невокализированном интервалов перекрываются даже на стационарных интервалах.

На рис. 6 представлены гистограммы распределений ЧПН и ЛЧПН вокализированных и невокализированных интервалов без добавления шума. По гистограммам видно, что диапазон возможных значений функции ЛЧПН значительно уже значений функции ЧПН. Гистограммы вокализированных и невокализированных интервалов в значительной степени перекрываются, причём область перекрытия для ЛЧПН меньше, чем для ЧПН.

Вероятность ошибки классификации для логарифмического масштаба частот пересечения нуля оказалась на 10–15% меньше, чем для линейного. Вероятность ошибки классификации Т/НТ по ЛЧПН для разных типов сигнала и различных дикторов изменялась в пределах 11%...21%. Добавление шума, подобно добавлению шума к признаку $R(1)$, несколько снижало ошибку классификации Т/НТ для чистого сигнала. Для других типов сигнала добавление шума практически не влияло на надёжность принятия решения Т/НТ. Выбирая уровень добавляемого шума при вычислении ЛЧПН, следует придерживаться тех

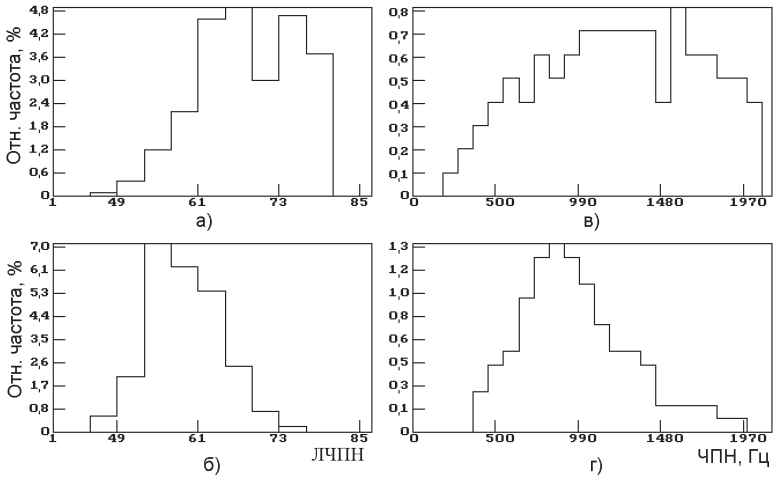


Рис. 6. Гистограммы распределений ЧПН и ЛЧПН:
а) невокализованные интервалы (ЛЧПН); б) вокализованные интервалы(ЛЧПН);
в) невокализированные интервалы (ЧПН); г) вокализированные интервалы (ЧПН)

же рекомендаций, что и при вычислении признака $R(1)$. В обоих случаях можно использовать единый генератор шума (для речи без искажений z=30 отс.; для телефонной речи z=15 отс.; для зашумлённой речи добавление шума нецелесообразно). Различия в поведении вероятности ошибки классификации были незначительны для синхронного с ОТ и асинхронного с ОТ способов вычисления признака ЛЧПН. С этой точки зрения, не имеет значения, каким способом вычислять ЛЧПН — синхронно или асинхронно с ОТ.

Принятие решения Т/НТ по совокупности признаков в многомерном пространстве признаков лишено наглядности представления распределений и требует больших вычислительных затрат, существенно большей обучающей выборки, а также процесса переобучения при изменении условий произнесения [11]. Для упрощения процедуры классификации решено объединить три указанных выше признака в один, исходя из следующих соображений. Коэффициент $R(1)$ и энергия в полосе частот имеют максимальные значения на тональных интервалах. ЛЧПН на тональных интервалах минимальна. Тогда обобщённый признак, по которому выполняют классификацию Т/НТ, может быть записан следующим образом:

$$G = \frac{R(1) \cdot E}{Z_{cr}}. \qquad (4)$$

На рис. 7 изображены осциллограмма фразы «Саша кусал сало» и соответствующий ей график обобщённого признака Т/НТ. Марки 2-11 установлены на границах вокализации. Обобщённый признак Т/НТ вокализованных звуков имеет большие значения по сравнению с признаком на невокализованных звуках.

В таблице 1 приведены значения вероятности ошибки классификации для обобщённого признака Т/НТ, а также для отдельных признаков классификации Т/НТ. Результатом объединения трёх признаков стало повышение точности классификации. Тем не менее, вероятность появления ошибки классификации остаётся достаточно высокой (см. табл. 1). Повышения точности распознавания можно достичь путём привлечения дополнитель-

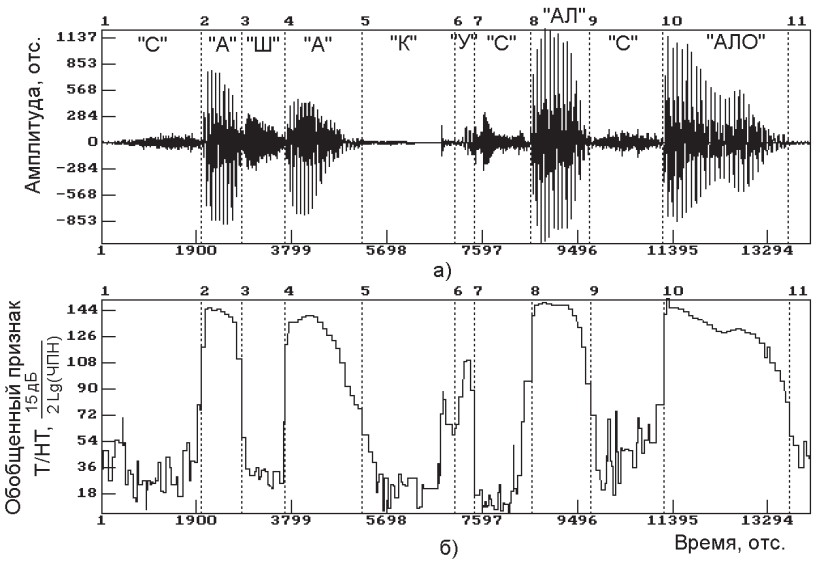


Рис. 7. Обобщённый признак T/HT речевого сигнала:
а) осциллограмма фразы «Саша кусал сало» (диктор — мужчина);
б) обобщённый признак T/HT сигнала

Таблица 1

Параметры классификации речи по коэффициенту R(1), энергии в полосе частот, ЛЧПН и обобщённому признаку для разных способов их вычисления

Способ вычисления признака	Признак классификации	Вероятность ошибки классификации
Чистый сигнал синхронно с ОТ	Коэффициент R(1)	0,0695
	Энергия	0,0735
	ЛЧПН	0,1135
	Обобщённый	0,059
Чистый сигнал асинхронно с ОТ	Коэффициент R(1)	0,098
	Энергия	0,104
	ЛЧПН	0,1125
	Обобщённый	0,104
Телефонный сигнал синхронно с ОТ	Коэффициент R(1)	0,1445
	Энергия	0,1465
	ЛЧПН	0,204
	Обобщённый	0,121
С/Ш 10 дБ синхронно с ОТ	Коэффициент R(1)	0,111
	Энергия	0,129
	ЛЧПН	0,175
	Обобщённый	0,101

ных признаков, определяемых предысторией процесса и длительностью интервалов, классифицированных как вокализированные или невокализированные [14].

В работе [19] при принятии решения Т/НТ с помощью порогов g_0, g_1 и g_2 область значений обобщённого признака разбивали на четыре области: «уверенно НЕ ТОН», «неуверенно НЕ ТОН», «неуверенно ТОН», «уверенно ТОН». Пороги g_0 и g_2 устанавливали так, что вероятности попадания вокализированного звука в невокализированную область и невокализированного звука в вокализированную не превышала 2%. При неопределённом решении о вокализации дополнительную информацию извлекали из априорных данных и известных значений длительностей предполагаемых периодов ОТ. Области «неуверенно НЕ ТОН», «неуверенно ТОН» относили к вокализированным или к невокализированным в ходе последующей обработки. Порог g_1, разделяющий области «неуверенно НЕ ТОН», «неуверенно ТОН», устанавливали из условия минимума вероятности суммарной ошибки классификации с учётом последующей обработки.

В таблице 2 представлены значения порогов классификации g_1, g_0 и g_2 для разных условий вычисления обобщённого признака. Значения порогов зависели от типа сигнала, а также от диктора и отдельных произнесений сигнала. Такая зависимость требует подстройки значений порогов для конкретных произнесений. Подобный способ установки порогов не способен учесть все возможные изменения произнесений и окружающей диктора обстановки. По этим причинам принято решение выполнять классификацию Т/НТ за два прохода.

Таблица 2

Значения порогов классификации

	Порог g0	Порог g1	Порог g2
Чистый сигнал синхронно с ОТ	67	76	97
Чистый сигнал асинхронно с ОТ	71	87	128
Телефонный синхронно с ОТ	48	65	143
С/Ш 10 дБ синхронно с ОТ	74	86	134

В предлагаемой модификации алгоритма на первом проходе вычисляют значение обобщённого признака G по формуле (4) для каждого периода ОТ. Эту процедуру выполняют как на вокализированных, так и на невокализированных участках речевого сигнала. На невокализированных участках сигнала за интервал анализа принимают интервал между двумя марками, проставленными алгоритмом *GS* [18] случайным образом. После окончания первого прохода для всего произнесения в целом строят гистограмму значений признака G (рис. 8) и вычисляют среднее значение признака G_t для данного произнесения. Эксперименты показывают, что величину G_t можно принять за первоначальную оценку границы между значениями обобщённого признака, соответствующими вокализированным ($G>G_t$) и невокализированным ($G<G_t$) интервалам речевого сигнала.

Для интервала значений $G<G_t$ (предположительно невокализированные звуки) вычисляют среднее значение обобщённого признака G_{uv} и среднеквадратическое отклонение σ_{uv}. Аналогично, для предположительно вокализированных звуков ($G>G_t$) вычисляют среднее значение обобщённого признака G_v и среднеквадратическое отклонение σ_v.

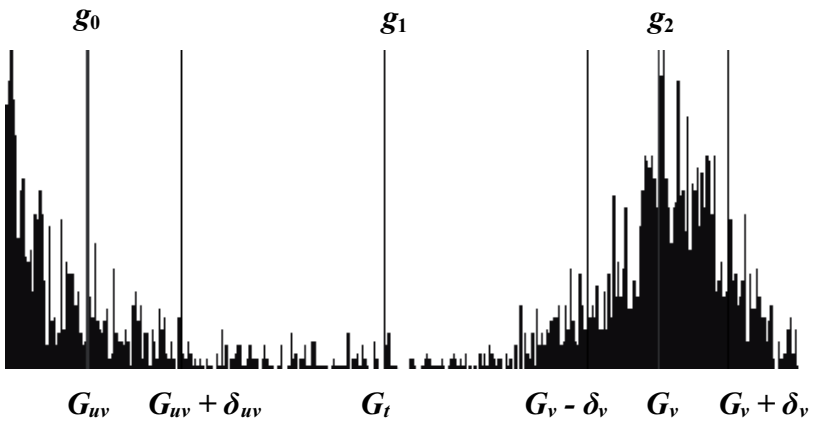


Рис. 8. Гистограмма обобщённого признака Т/НТ всех произнесений диктора АИО

Таблица 3

Значение обобщённой ошибки при различных значениях g0, g1, g2

g_0	g_1	g_2	Обобщённая ошибка
G_{uv}	G_t	G_v	2,76
G_{uv}	G_t	$G_v+\delta_v$	3,13
G_{uv}	G_t	$G_v - \delta_v$	3,08
$G_{uv}+\delta_{uv}$	G_t	G_v	4,50
$G_{uv}- \delta_{uv}$	G_t	G_v	3,54
$G_{uv} - \delta_{uv}$	G_t	$G_v-\delta_v$	3,50
$G_{uv}- \delta_{uv}$	G_t	$G_v + \delta_v$	3,57
$G_{uv} + \delta_{uv}$	G_t	$G_v - \delta_v$	4,47
$G_{uv}+\delta_{uv}$	G_t	$G_v+\delta_v$	4,57

Исследовано несколько экспериментальных правил задания значений порогов g_0, g_1, g_2. Эти правила сведены в таблицу 3. В этой же таблице показаны значения обобщённой ошибки (ОШ), получаемые двухпроходным алгоритмом для каждого из выбранных правил задания порогов. Обобщённая ошибка учитывает значения ошибок «ТОН–НЕ ТОН», ошибок «НЕ ТОН» и больших ошибок, оцениваемых путём сравнения измеренного контура ОТ с эталонным по правилу, изложенному в работе [24]. Из таблицы 3 следует, что минимальная обобщённая ошибка (ОШ=2,76%) получена в том случае, когда значения порогов g_0, g_1, g_2 устанавливали из соотношений:

$$\begin{cases} g_0 = G_{uv} \\ g_1 = G_t \\ g_2 = G_v \end{cases}. \qquad (5)$$

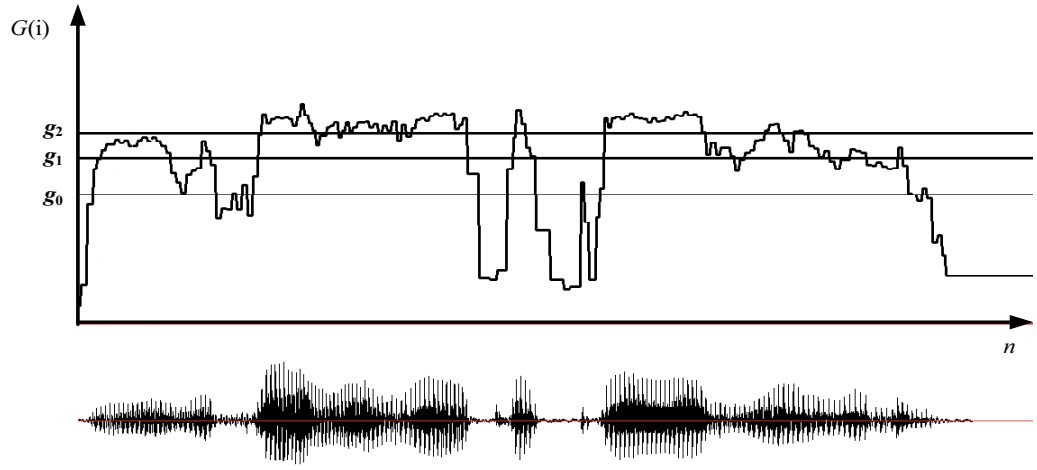


Рис. 9. Речевой сигнал (внизу) и обобщённый признак Т/НТ G(i) (вверху) с отображёнными порогами g0, g1, g2 для диктора AIO (фраза: «Не видали мы такого невода»)

На рис. 9 представлен пример осциллограммы предложения «Не видали мы такого невода» (диктор AIO); траектория обобщённого признака $G(i)$ (i — порядковый номер периода ОТ) для данного произнесения и значения порогов g_0, g_1, g_2, выбранные по правилу (5).

Окончательные решения Т/НТ получали путём коррекции предварительных решений «уверенно ТОН», «уверенно НЕ ТОН», «неуверенно ТОН», «неуверенно НЕ ТОН». При окончательном решении Т/НТ по предварительной оценке «неуверенно ТОН», «неуверенно НЕ ТОН», учитывали относительную нестабильность соседних периодов ОТ. Вокализированные участки длительностью меньше 20 мс относили к невокализированным.

В таблице 4 представлены результаты сопоставительных испытаний двухпроходного алгоритма классификации Т/НТ, совмещённого с выделителем ОТ, по методу *GS* с шестью выделителями ОТ и признака Т/НТ, реализованных в системе SIS [23]: с пиковым, фильтровым, автокорреляционным, кепстральным методами, с методом Голда-Рабинера и методом ЛЛК.

Таблица 4

Результаты испытаний алгоритмов выделения ОТ для общей группы голосов (15 дикторов, 38 произнесений)

Выделитель ОТ	Ошибка ТНТ, %	Ошибка НТТ, %	ТНТср %	Большие ошибки, %	Малые ошибки, %	Обоб. ошибка, %	Отношение ТНТср/ ОШ
Чистый сигнал							
GS2	1.97	2.37	2.17	1.70	6.16	2.76	0.79
Пиковый	0.62	7.27	3.94	1.23	10.06	4.13	0.95
Кепстральный	3.89	5.50	4.70	3.76	19.95	6.02	0.78
АКФ	1.67	21.44	11.55	2.98	10.00	11.93	0.97

Рабинер-Гоулд	1.93	8.52	5.23	2.87	9.59	5.96	0.88
Фильтровой	0.11	14.83	7.47	1.16	9.02	7.56	0.99
ЛЛК	0.64	6.29	3.47	1.08	6.63	3.63	0.95
Сигнал с аддитивным шумом С/Ш = 5 дБ							
GS2	2.01	18.18	10.10	15.58	20.72	18.56	0.54
Пиковый	1.47	36.22	18.84	7.03	24.26	20.11	0.94
Кепстральный	1.95	37.90	19.92	11.85	36.75	23.18	0.86
АКФ	2.63	36.30	19.46	3.35	13.56	19.75	0.99
Рабинер-Голда	2.10	32.57	17.33	4.00	17.23	17.79	0.97
Фильтровой	1.38	41.44	21.41	2.81	22.55	21.59	0.99
ЛЛК	1.01	45.93	23.47	2.13	20.61	23.57	1.00
Сигнал ограничен полосой телефонного канала							
GS2	0.61	14.74	7.67	6.19	6.38	9.86	0.78
Пиковый	0.90	19.50	10.20	12.08	9.89	15.81	0.65
Кепстральный	4.98	15.29	10.14	2.91	19.03	10.55	0.96
АКФ	1.06	46.60	23.83	5.36	11.06	24.42	0.98
Рабинер-Голда	2.37	19.00	10.69	28.76	5.22	30.68	0.35
Фильтровой	0.10	37.56	18.83	10.10	6.81	21.37	0.88
ЛЛК	0.10	37.56	18.83	10.10	6.81	21.37	0.88

Литература

1. Сапожков М.А. Речевой сигнал в кибернетике и связи. М.: Связьиздат, 1963. 472 с.

2. Гитлин В.Б. Основной тон речевого сигнала / Деп. В ВИНИТИ, 1998. №1206-В98. 739 с.

3. Сапожков М.А., Михайлов В.Г. Вокодерная связь М.: Радио и связь, 1983. 248 с.

4. Вокодерная телефония / Под ред. Пирогова А.А. М.: Связь, 1974. 536 с.

5. Miller N.J. Pitch detection by data reduction // IEEE Symp. speech recogn. Carnague-Mellon Univ., 1974. Contrubut Pap. P.122–130.

6. Friedman D.H. Multidimensional Pseudo-Maximum Likelihood pitch estimation // IEEE Trans. Acoust., Speech and Signal Process. 1978. Vol.26. N3. P.185–196.

7. Маркел Дж. Д., Грэй А.Х. Линейное предсказание речи. М.: Связь, 1980. 308 с.

8. De Souza P. A statistical approach to the design of an adaptive self-normalising silence detector / IEEE Trans. Acoust., Speech and Signal Process. 1983. 31. N3. P.678–684.

9. Foo S.W., and Turner L.F. Application of sub-band energy ratio to Voiced-Unvoiced-Silence classification of speech signals // Proc. MELECON'83 Mediterr. Electrotechn.Conf. Athens, 24-26, May, 1983, Vol. 2. S1. Sa. 1983. C3.05/1 — C3.05/2.

10. Atal B.S. Speech signal pitch detector using prediction error date. Pat. N 3740476 USA. G10L 1/04. 19.06.73.

11. Atal B.S., Rabiner L.R. A pattern recognition approach to voiced-unvoiced-silence classification with application to speech recognition // IEEE Trans. Acoust., Speech and Signal Process. 1976. 24. N3. P.201–202.

12. Hebid M.K., and Robinson D.M., Sincoscie W.D. Real Zeros in pitch detection // IEEE Int. Conf. Acoust., Speech and Signal Process. Record. Tulsa, Okla, 1978. New York, N.Y. 1978. P.31–34.

13. Кельманов А.В. Алгоритм классификации тон/шум, основанный на критерии адекватности модели авторегрессии // Вычислительные системы. Методы обработки информации. Новосибирск, 1978. Вып.74. С. 129–148.

14. Кельманов А.В. Алгоритм классификации тон/шум по частотным автокорреляциям // Вычислительные системы. Эмпирическое предсказание и распознавание образов. Новосибирск, 1980. Вып.83. С. 67–73.

15. Рабинер Л.Р., Шафер Р.В. Цифровая обработка речевых сигналов. М.: Радио и связь, 1981. 485 с.

16. Дуда Р., Харт П. Распознавание образов и анализ сцен. М.: Мир, 1976. 512 с.

17. Баронин С.П. Автокорреляционный метод выделения основного тона речи // Сб. тр. Гос. НИИ Министерства связи СССР. 1961. 3(24). С. 93–102.

18. Архипов И.О., Гитлин В.Б. Метод выделения основного тона на основе понятия о генерируемом солитоне // Распознавание образов и анализ изображений: новые информационные технологии. 4-я Всероссийская с международным участием конференция. РОАИ-98. 1998 г. Новосибирск, 1998. Часть 1. С. 23–27.

19. Архипов И.О., Гитлин В.Б. Формирование признака ТОН/НЕ_ТОН синхронно с основным тоном // Современные речевые технологии. Сборник трудов IX сессии Российского акустического общества. М.: ГЕОС, 1999. С. 43–46.

20. Архипов И.О., Гитлин В.Б. Добавление шума при сегментации речи на тональные участки // Труды научно-молодёжной школы «Информационно-измерительные системы на базе наукоёмких технологий». Ижевск: изд. ИПМ УрО РАН, 1997. с. 63–69.

21. Архипов И.О., Гитлин В.Б. Сегментация речи по первому коэффициенту линейного предсказания синхронно с основным тоном // Труды научно-молодёжной школы «Информационно-измерительные системы на базе наукоёмких технологий». Ижевск, изд. ИПМ УрО РАН, 1998. С. 17–19.

22. Архипов И.О., Гитлин В.Б. Оценка частоты среза ФНЧ, используемого для выделения основного тона // Труды научно-молодёжной школы «Информационно-измерительные системы на базе наукоёмких технологий». Ижевск: изд. ИПМ УрО РАН, 1998. С. 12–16.

23. Методические рекомендации по практическому использованию программы SIS при работе с речевыми сигналами / Центр речевых технологий. Санкт-Петербург, 1997. 394 с.

24. Архипов И.О., Гитлин В.Б. Оценка точности выделения основного тона методом GS // Современные речевые технологии. Сборник трудов IX сессии Российского акустического общества. М.: ГЕОС, 1999. С. 38–42.

Архипов Игорь Олегович,
кандидат технических наук, доцент кафедры «Программное обеспечение ЭВМ» Ижевского технического университета (426069, Ижевск, ул. Студенческая, 7).

Гитлин Валерий Борисович,
доктор технических наук, профессор кафедры «Вычислительная техника» Ижевского технического университета. E-mail: vbg_istu@mail.ru, vbg@mitm.ru.

Лузин Дмитрий Александрович,
аспирант кафедры «Вычислительная техника» Ижевского технического университета.

О допустимых пределах искажений электроакустических речевых сигналов при скрытом встраивании данных

М.О. Пономарь

Предлагаемый метод сокрытия данных в речевых сигналах основан на использовании стеганографии под прикрытием поточной криптозащиты.

Особый интерес для систем скрытой связи по открытым речевым каналам представляют те методы, в которых скрываемые данные внедряются в значения непрерывных несущих параметров: время запаздывания эхо-сигнала, значения фазы спектральной составляющей, значения частоты основного тона и длительности вокализованных сегментов речи. При этом скрываемые данные оказываются достаточно стойкими к воздействию шумов, фильтрованию, сжатию с потерями, вокодерному, аналого-цифровому, цифро-аналоговому преобразованиям и для их извлечения не требуется исходный аудиосигнал [1].

При внедрении дискретных данных в непрерывные характеристики речевого сигнала требуется использовать искусственное квантование его по времени и уровню. Сегментация сигнала на естественные однородные вокализованные стационарные участки является аналогом его квантования по времени, а для квантования значений несущих параметров по уровню наиболее простым в поточной реализации является метод кодирования с модуляцией индекса квантования (Quantization Index Modulatin — QIM) [2].

Физически результат преобразования кодером QIM, например, частоты основного тона (ЧОТ) на передающей стороне, состоит в том, что из естественных, произвольных по частоте сегментов речи, поступающих на вход кодера, на выходе кодера формируются сегменты речи с нормированными частотами, соответствующие центрам интервалов квантования. На приёмном конце канала связи декодер извлекает из них скрытые данные на основе определения значений принятых ЧОТ сегментов и сопоставления их с общей для передающей и приёмной сторон кодовой таблицей [3].

Совершенно очевидно, что нарушитель, обнаружив в речи абонента нормированные частоты (нормированные значения эхо-сигнала, фазы или дли-

Рис. 1. Схема защиты скрываемых данных путём встраивания их с помощью искусственной модификации речи QIM-кодером и передачи вместе с речью в открытый канал под прикрытием криптографического преобразования

тельности вокализованных сегментов речи), легко определяет наличие в канале связи скрытых данных. Он даже сможет сразу прочесть их, в случае если они представляют собой сообщение и переданы открытым текстом в какой-либо из известных ему кодировок. В случае если сообщение зашифровано, в дело вступает криптоаналитик. Как известно, поточные шифры значительно менее стойки к дешифрации, чем блочные, а это значит, что есть шансы и у криптоаналитика. Но главное то, что стегоканал им обнаружен. В зависимости от результатов криптоанализа и цели нарушителя он может продолжать прослушивать канал или воздействовать на него, например, с целью разрушения скрытого сообщения или навязывания получателю ложной информации — то есть превратиться в активного нарушителя.

Из этого следует, что проектируемый стегоканал должен быть защищён от обнаружения и единственным практическим способом его защиты является криптозащита. В данном методе сокрытия демаскирующим признаком скрываемого сообщения являются нормированные значения несущих параметров, то есть необычные, неестественные статистические свойства заполненного контейнера по сравнению с пустым. Можно применить дизеринг — добавление небольшого шумового сигнала, делающего основной сигнал более естественным, но это затруднит его декодирование.

Единственным радикальным решением является криптозащита. Это значит, что на выходе кодера QIM необходимо иметь криптографический преобразователь, который формирует из нормированных параметров новые параметры, похожие на произвольные, естественные для человеческой речи, но зашифрованные. На приёмном конце происходит сначала расшифровывание каждого параметра, а затем QIM-декодирование его с целью извлечения скрытых данных. Нарушитель при этом не получает никаких сведений о наличии скрытого сообщения, а тем более не может его прочитать, так как он прослушивает полностью естественную речь, часть параметров которой

Таблица 1

Пример встраивания данных с использованием скрывающей модификации частоты основного тона сегментов речи с их криптозащитой

№ сегмента	1	2	3	4	5	6	7	8	9
ЧОТ пустого контейнера, container	114	114	114	130	206	115	115	159	206
Стего-коды пустого контейнера	14	14	14	30	06	15	15	59	06
Стего-вложение (симв/ дв/дес)	**рус/** 00000/ 0	**К/** НПО/ 30	**О/** 00011/ 03	**М/** 00111/ 39	**П/** 01101/ 13	**А/** 11000/ 24	**Н/** 00110/ 06	**И/** 01100/ 44	**Я/** 1110/ 29
ЧОТ заполненного стегоконтейнера, stego	100	130	103	139	213	124	106	144	229
Гамма (симв/дв)	**Лат./** 11111	**Е/** 10000	**М/** 00111	**Б/** 10011	**Е/** 10000	**Б/** 10010	**I/** 01100	**N/** 00110	**C/** 01011
Шифро-текст (симв/дв/ дес)	**Лат./** 11111/ 31	**C/** 01110/ 14	**Проб/** 00100/ 04	**S/** 10100/ 20	**Q/** 11101/ 18	**R/** 01010/ 10	**R/** 01010/ 10	**R/** 01010/ 10	**F/** 1011/ 22
ЧОТ заполненного крипто-стегоконтейнера, stego + crypto	131	114	104	120	218	110	110	110	222

модифицирована в пределах психоакустической нормы с использованием криптозащиты.

Необходимо подчеркнуть, что это достаточно сложная задача, поскольку слуховое восприятие настолько совершенно, что позволяет опознать самые тонкие оттенки речевого сигнала. Человеческий слух, а тем более слух акустического стегоаналитика довольно точно определит признаки искусственности и естественности речи. И при встраивании данных необходимо учитывать два фактора: неслучайность характера сигналов незаполненного речевого контейнера и сохранение его качества при встраивании и шифровании данных.

Таким образом, речевой сигнал при встраивании в него данных и их извлечении должен претерпевать два прямых и два обратных стего- и криптопреобразования. Покажем, что эти преобразования алгоритмически реализуемы.

Воспользуемся для этого примером встраивания данных в модификацию ЧОТ, приведённым в работе [1]. Речевой контейнер со словами «Wow… Sound editing just…» длительностью 2 сек. разделён на 9 участков с приведёнными в первой строке таблицы 1. ЧОТ (в целых числах Гц, плотность вложения — 1 буква на сегмент речи в гомофоническом коде типа МТК-2 , интервал стегодекодера от –0,5 Гц до +05 Гц).

Достаточно длительные незашифрованные последовательности, подобные приведённому в строке 3 встроенному слову КОМПАНИЯ, будут легко обна-

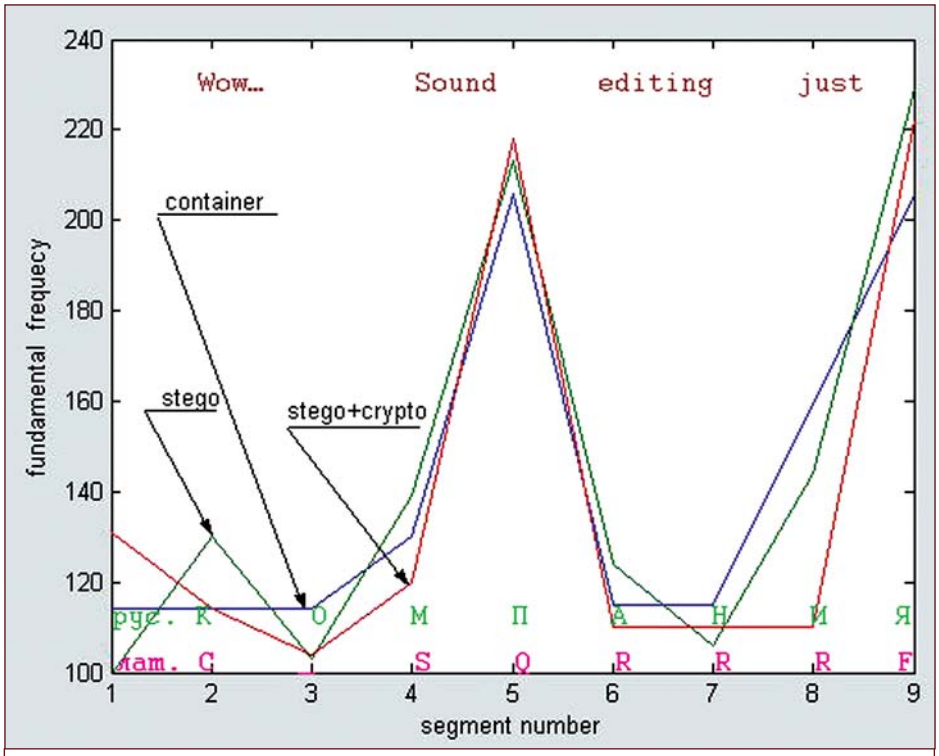


Рис. 2. Пример встраивания данных в речь путём стего- и криптопреобразования частоты основного тона сегментов речи.

ружены и прочтены нарушителем путём статистического анализа заполненного контейнера. Произведём шифрование этого слова. В поточных шифраторах каждый бит исходной информации шифруется с помощью гаммирования — наложения обратимым образом на открытые данные последовательности псевдослучайных чисел. В данном примере в качестве гаммы использовано слово EMBEDING с наложением побитовым «исключающим ИЛИ» (XOR). Получившийся шифротекст C_SQRRRF уже менее доступен криптоаналитику для прочтения, однако шифрование может повлиять на качество речевого контейнера (рис. 2), а значит, вызвать подозрение о наличии в нём стеговложения.

Но, как видно из рис. 2, в данном случае шифрование не вызвало деградирующих преобразований ЧОТ речи, сохранены и мелодический контур, и интонация фразы. Это объясняется тем, что используемый стегокод имеет диапазон значений от 0 до 99, а накладываемая гамма в коде МТК-2 только от 0 до 32. Можно назвать это криптографическим поточным преобразованием, сходным с дизерингом. Вполне очевидны метод расшифрования повторным наложением гаммы на текст и метод извлечения данных с использованием кодовой таблицы. Однако вопрос о допустимом пределе модификации ЧОТ и других параметров речи с учётом шифрования требует дальнейшего исследования.

Решение задачи встраивания данных в речь с шифрованием также тесно связано с непростой задачей синхронизации потокового криптопреобразования, которое требуется как при начале скрытой передачи данных, так и при её продолжении в случае временной потери связи. Трудность обеспечения синхронизации, по мнению некоторых авторов,

превращается в достоинство с точки зрения обеспечения скрытности передачи [4], поэтому использование искусственных средств синхронизации — синхронизирующих посылок, меток, заголовков и т.п. — является крайне нежелательным. Необходима оценка технической сложности задачи синхронизации и возможности её решения в реальном масштабе времени.

Заключение

В данной работе не рассматривались вопросы реализации и криптографической стойкости поточных шифров, это является предметом исследования других специалистов. Однако использование стегокодирования в сочетании с шифрованием вносит дополнительные искажения в речевой сигнал и задержки при его передаче. В связи с этим при разработке технических средств скрытной передачи данных потребуется определить их допустимые пределы, как с точки зрения скрытности, так и с точки зрения возможности поддержания речевой коммуникации.

Литература

1. Ponomar, Marina. Data hiding in speech signals on the basis of the modification of segment pitch and duration. 19th International Congress on Acoustics ICA2007MADRID, 2–7 Sept. 2007, Madrid, Spain, 2007, CAS-03-023, p.46.

2. Chen B., Wornel G.W. System, method, and product for information embedding using an ensemble of non-intersecting embedding generators. U.S. patent pending. Licensing info.: MIT Technology Lic. Office. 1996.

3. M.O.Ponomar. Coding with the Quantization of Speech Signal Carrier Features for Data Hiding. XX Session of the Russian Acoustic Sociaety. T.3. M.: GEOS, 2008, p.645–648.

4. Грибунин В.Г., Оков И.Н., Туринцев И.В. Цифровая стеганография. «Методы и технические средства обеспечения безопасности информации». — СПб.: ГТУ, 2001.

Пономарь Марина Олеговна,
аспирантка кафедры прикладной и экспериментальной лингвистики Московского государственного лингвистического университета
E-mail: oponomar@inbox.ru.

От звучащей речи — к жестовой

А.Л. Воскресенский
Г.К. Хахалин

В статье даётся предварительное описание подхода к созданию системы автоматизированного сурдоперевода. Приводится обоснование необходимости в создании такой системы, описание некоторых её особенностей, возможных путей разрешения проблем.

Введение

Необходимость в создании систем автоматизированного сурдоперевода диктуется не только требованиями мирового сообщества по обеспечению равных прав для всех [1]. Не все глухие в достаточной степени понимают текст сообщений на информационных табло с выводом текста «бегущей строкой», что связано с меньшим (по сравнению со слышащими) объёмом их активного словаря.

В данной работе представляются как результаты исследований, проводившихся в течение ряда лет, так и описание подхода к решению возникающих в ходе работы задач. В отличие от предшествующих публикаций (например, [2, 3]), основное внимание уделяется проблемам поиска конкретных значений понятий, возникающих при переводе звучащей речи в жесты. При этом используются примеры, зафиксированные при разработке толкового словаря русского жестового языка RuSLED [4, 5].

Основные сложности при переводе текста (который может быть результатом работы подсистемы распознавания звучащей речи) связаны с разрешением омонимии, нахождением необходимого значения полисемичного слова, а также с преобразованием фраз русского языка, имеющих свободный порядок слов, в жестовые выражения, в которых порядок жестов значительно более строг.

По своим функциям и характеристикам система перевода текста в жесты может быть отнесена к системам искусственного интеллекта [6], при этом решаемые задачи поиска требуемого значения слова или совокупности слов в некоторых случаях сложнее, чем при переводе с одного словесного языка на другой.

1. Краткое описание словаря RuSLED

Словарь русского жестового языка RuSLED (Russian Sign Language Explanatory Dictionary) включает в себя функции толкового словаря как для введённого слова, так и для его жестового представления. На вход словаря подаётся произвольная форма слова, а на выходе демонстрируются варианты жестового толкования данного слова.

Словарь содержит 2372 слова (с толкованиями их значений) и 2537 видеоизображений жестов (включая различные варианты исполнения), передающих значения этих слов. Для 1592 жестов (63% от общего числа, вошедших в словарь) даны дополнительные пояснения, относящиеся к манере исполнения жеста или описывающие смысловые нюансы, передаваемые жестом.

В словаре представлены жесты, используемые в Санкт-Петербурге и его окрестностях. Частично представленные в словаре жесты совпадают с жестами, используемыми в Москве, но в целом расхождения достаточно велики, что дало повод назвать данный словарь «Петербургский диалект».

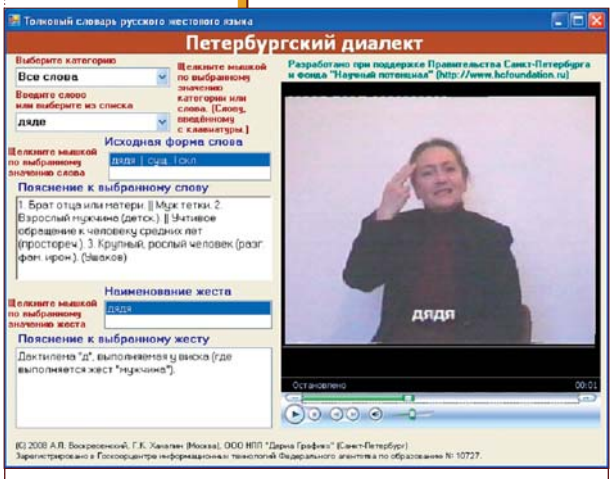


Рис. 1. Экранная форма словаря RuSLED

Видеоряд словаря составлен на основе видеокурса, изданного Межрегиональным центром реабилитации (МЦР), г. Павловск [7]. В данной версии словаря для демонстрации жестов используются оцифрованные фрагменты видеозаписи сурдопереводчиков, заимствованные из видеокурса. Использование для просмотра жеста элемента ActiveX Windows Media Player позволяет:

- просмотреть этот же жест повторно при нажатии кнопки плеера },
- приостановить выполнение жеста в требуемом месте при нажатии кнопки плеера II,
- просмотреть любую фазу выполнения жеста, передвинув мышью движок плеера в соответствующую позицию (рис. 1).

Поставленная ранее цель — использование для демонстрации жестов виртуального персонажа (аватара) — пока не достигнута из-за сложности представления мимики, сопровождающей жесты и выполняющей весьма важную роль в жестовом языке глухих. Так, например, слова *милый*, *симпатичный* передаются одним жестом, но отличаются движениями губ, проговаривающих фрагменты соответствующих слов.

При составлении пояснений к некоторым жестам использовались пояснения из словаря «Говорящие руки» Фрадкиной [8], составленного на основе московского варианта жестового языка.

При составлении пояснений к словам использованы более 30 словарей и энциклопедий, доступ к которым осуществлялся через Интернет, с использованием, по большей части, службы «Словари» портала Яндекс, за исключе-

нием нескольких словарей — в частности, одной из версий Толкового словаря русского языка Ушакова, размещенной на портале ГРАМОТА.РУ.

По рекомендациям сурдопедагогов, обеспечена возможность фильтрации словника словаря по грамматическим категориям (существительные, глаголы, прилагательные, наречия, предлоги, частицы, числительные, местоимения). Для просмотра всего содержимого словаря нужно выбрать категорию «Все слова».

Программная оболочка словаря зарегистрирована Госкоорцентром информационных технологий (ОФАП Минообразования и науки РФ) №10727 от 30.05.2008.

Дистрибутив словаря на DVD выполнен и распространяется ООО НПП «Дериа Графикс» (г. Санкт-Петербург).

Отличием словаря является то, что для каждого семантического значения лексемы (и жеста) используется отдельный вход словаря — отдельная запись в таблице базы данных. Это значительно удобнее для пользователя, является очевидным решением для электронных толковых словарей и рекомендуется лексикографами [9].

Поле «Введите слово» позволяет вводить произвольные словоформы или выбирать из списка лексемы, имеющиеся в словаре. В список «Исходная форма слова» выводится соответствующее основе значение лексемы или несколько значений, если по результатам морфологического анализа выбрано несколько записей.

При выборе пользователем нужной лексемы в поле «Наименование жеста» выводится наименование жеста (как правило, совпадающее с лексемой) или (если данной лексеме соответствуют несколько жестов) список наименований жестов. Для каждого из значений слова выдаётся только то значение жеста, семантика которого соответствует значению выбранного из списка слова [4].

2. Примеры неоднозначности слов и соответствующие процедуры обработки контекста

Поскольку между жестами и словами нет однозначного соответствия, при переводе текста в жесты необходимо не только разрешать проблемы омонимии (которые в ряде случаев могут быть сняты лингвистическими средствами путём анализа морфологических форм слов и синтаксиса фраз, в которых они встречаются), но и осуществлять тщательный отбор нужного значения полисемичного слова из соответствующего ряда синонимов.

Наблюдения сурдопедагогов [10] показывают, что абстрактно-логический уровень мышления у глухих формируется позднее, чем у слышащих. В результате у глухих превалирует предметно-образный уровень мышления. Поэтому, как показано ниже, в ряде случаев использование синонима вместо точного значения допустимо при переводе с одного словесного языка на другой (слушающий подсознательно подставляет вместо услышанного слова нужное значение), тогда как при переводе на жестовый язык мы должны найти точное значение слова, иначе мы не сможем подставить в формируемое жестовое выражение нужный жест.

Использованные ниже примеры основаны на словах, имеющихся в словаре RuSLED.

2.1. Омография некоторых форм слов

В русском языке написания слова *вино* в родительном падеже единственного числа и слова *вина* в именительном падеже совпадают. Эти примеры могут быть продолжены: например, совпадают написания существительного *весть* в именительном падеже множественного числа и родительном падеже единственного числа, а также глагола *вести*.

Здесь для выявления нужного значения слова достаточно использовать синтаксический анализ локального контекста (ближайшего окружения слова, зачастую меньшего, чем предложение в целом), позволяющий выбрать нужную лексему из вариантов, предлагаемых морфологическим анализатором. При этом учитываются согласованность прилагательных и существительных и связность предложения, включающего анализируемые цепочки слов [11].

2.2. Некоторые случаи омонимии

Приведём несколько примеров. Так, словом *лук* в русском языке обозначаются как съедобное растение, так и вид метательного оружия; словом *автомат* обозначаются как вид огнестрельного оружия, так и устройство, работающее по заданной программе.

Здесь для выявления нужного значения подчёркнутого слова уже не достаточно использовать синтаксический анализ локального контекста. Необходимо использовать контекст, выходящий за пределы предложения [12]. При этом необходимо учитывать частотные характеристики встречаемости слов в рассматриваемом контексте [13, 14], не исключая из рассмотрения предлоги [14], которые часто относятся к категории «стоп-слов», не учитываемых при анализе. Таким образом, помимо достаточно обширного словаря и знания грамматики, система обработки текста должна иметь примеры употребления слов, входящих в её словарь, имеющие ссылки на соответствующие семантические классы.

2.3. Полисемия

Слово *земля* в русском языке имеет ряд значений, из которых в словаре RuSLED встречаются значения *планета, почва, берег*. Рассмотрим последний случай (рис. 2).

Для жеста, передающего значение *берег*, в словаре [8] приводится пояснение: *«"Земля!" — закричали матросы»*. Различные программы-переводчики, доступные в Интернете, дают следующие варианты перевода (примеры a, b, c):
(a) *«Ground!» — sailors cried (Cognitive Translator, http://cs.isa.ru:10000/ct/);*
(b) *«The Earth!» — sailors have cried (PROMT© Translator, http://www.translate.ru/);*
(c) *«Land!» — cried the sailors (Translator Google©, http://translate.google.com/).*

Общаясь с помощью словесной речи, мы каждый раз решаем задачу распознавания информации, передаваемой нам собеседником. При этом происходит

подстановка значений слов, хранящихся в нашей памяти, т.е. воспринятый смысл текста не является точным аналогом слов, составляющих фразы текста. Там, где это возможно, воспринятое содержание фразы внутренне дополняется (и корректируется) в соответствии с общим содержанием текста и имеющимися знаниями об окружающем мире, не вызывая проявляемых внешне затруднений и протеста. Поэтому варианты (a) и (c) могут быть признаны допустимыми для случая словесного языка, а вариант (b) — нет, поскольку «The Earth» означает планету Земля, которую матросы не могут увидеть как цельный объект ни при каких обстоятельствах.

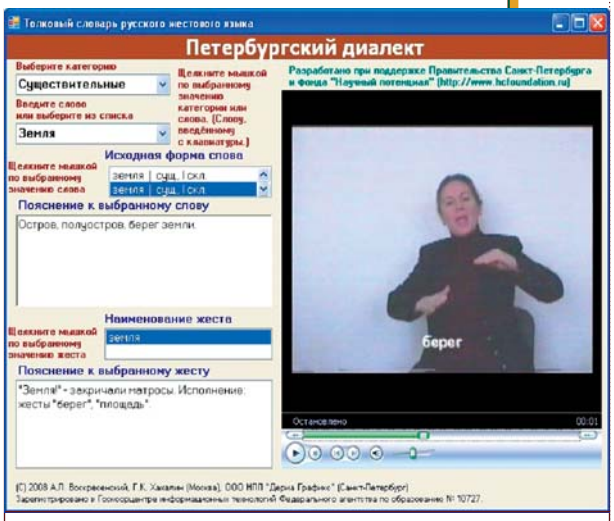


Рис. 2. Жест «берег» в ряду синонимов слова «земля» в словаре RuSLED

Но отметим, что ни в одном из случаев не получено значение coast (берег), необходимое для задачи сурдоперевода. То есть система сурдоперевода должна самостоятельно решать задачу выбора и подстановки нужных значений слов, исходя из общего содержания текста. Эти значения не всегда, как показывают приведённые примеры (a)–(c), будут очевидными, поэтому такая задача с полным основанием может считаться интеллектуальной.

Поясним ход рассуждений системы в данном случае, приводящих к распознаванию ситуации [15]: матросы находятся на корабле, пребывающем в открытом море → корабль со всех сторон окружён водой → граница воды и суши (земли) называется *берег* → если матросы закричали «Земля!», это означает, что они увидели границу между водой и сушей (землёй), т.е. *берег (coast)* (рис. 3).

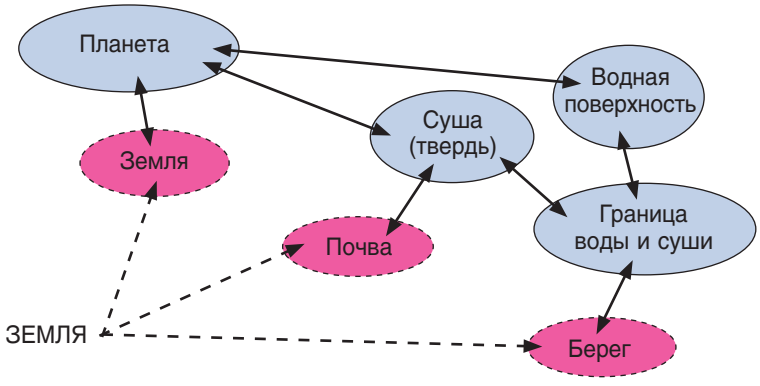


Рис. 3. Упрощённая схема выбора синонимов слова «земля»

Представленные рассуждения соответствуют традиционной системе логических умозаключений, известной со времён Аристотеля. Из известных прототипов систем искусственного интеллекта, использующих подобный подход, можно назвать, например, системы NARS и Novamente [16].

Для выполнения подобных рассуждений система должна иметь обширные знания об окружающем мире (или, по крайней мере, о тематике обрабатываемых текстов), форми-

рующие внутреннюю онтологию системы [6]. Эти знания могут пополняться не только за счёт содержания обрабатываемых текстов, но и из внешних источников, например, из сети Интернет.

При этом система должна «иметь собственное мнение» об окружающем мире и значениях слов, описывающих этот мир, поскольку семантическая разметка внешних источников информации может не отвечать требованиям решаемой задачи.

3. Возможный путь повышения качества распознавания речи

Система автоматизированного сурдоперевода, чтобы действительно быть полезной (например, во время публичных выступлений, лекций и т.п.), должна иметь подсистему распознавания речи. Соответственно, качество сурдоперевода будет во многом зависеть и от качества распознавания речи на входе системы.

Отвлечёмся от значений качества распознавания речи имеющихся систем и прототипов. Несомненно, что ко времени создания средств перевода текста в жесты они возрастут. Но чем ближе качество распознавания будет приближаться к идеалу, тем труднее будет движение к нему, тем больше затрат будет приходиться на каждую долю процента повышения качества работы системы распознавания.

Похожая ситуация была с системами оптического распознавания символов (OCR), которые используются для ввода в компьютер текста. В начале 90-х годов на российском рынке было представлено довольно много подобных систем, конкурирующих между собой. В рамках выполнения работ по созданию каталога российских коллективов, работающих в области обработки текста и речи [17], была поставлена задача оценки различных систем OCR. Практически все разработчики давали сведения о точности распознавания, равной 95–97%.

Но распознанный системой OCR текст ещё не является конечным продуктом. Для окончательной обработки необходимо выявить и устранить ошибки распознавания. Частично такая обработка выполняется путём автоматической проверки полученного текста с помощью словаря, включённого в систему OCR. Оставшиеся ошибки (и добавленные за счёт ошибочных срабатываний системы орфокоррекции) исправляются вручную. В [18] была предложена методика выявления ошибок распознавания, позволившая определить наиболее эффективные системы. Позднее данный прогноз оправдался.

Поскольку результаты работы системы распознавания речи во многих случаях должны представлять текст, целесообразно для коррекции ошибок распознавания использовать методы обработки полученного текста (морфологические, синтаксические, семантические).

Заключение

Работа по созданию системы автоматизированного сурдоперевода находится в начальной стадии. Имеющиеся предварительные результаты показывают, что такая система должна не просто конвертировать речь (и текст) из одной формы в другую, но и вести интеллектуальную обработку. Фактически система должна понимать обрабатываемые сообщения.

Таким образом, эта система должна решать задачи не менее, а возможно, и более сложные, чем системы автоматизированного перевода с одного словесного языка на другой.

Мы надеемся, что успешное завершение начатой работы окажется полезным не только для нужд сурдоперевода, но и для более широкого круга применений.

Литература

1. UN Resolution A/RES/48-96 (Part II, Rule 5, paragraph 7). http://www.un.org/documents/ga/res/48/a48r096.htm.

2. Voskresenski A. «Computer bank of sign languages», Conference abstracts, WISTCIS Outlook Conference «Information Society Priorities: New Prospects for European CIS Countries». Moscow, Russia, 20–21 November, 2003.

3. Voskresenski A. Signs and speech: two forms of human communication.. // Proceedings of the Ninth International Conference «Speech and Computer» SPECOM'2004, Saint-Petersburg, Russia, 2004. P. 666–669.

4. Voskresenskij A., Khakhalin G. Semantic Search Engine in a Multimedia Russian Sign Language Dictionary // Proceedings of the XIIth International Conference «Speech and Computer» SPECOM'2007, October 15–18, 2007, Moscow, Russia. P. 739–744.

5. Voskresenskiy A.L., Gulenko I.E., Khakhalin G.K. RuSLED Dictionary as Tool for Semantic Study // Proceedings of XV International Conference «Dialogue-2009» on Computer Linguistics and Intellectuall Technologies, May 27–31, 2009, Moscow, Russia.

6. Voskresenskij A. Text Disambiguation by Educable AI System.. // The First Conference on Artificial General Intelligence / P. Wang et al. (Eds.), AGI-08, 1–3 March, 2008, Memphis, USA. IOS Press, 2008.

7. Специфические средства общения глухих. Видеокурс. В 3 частях // СПб — Павловск: МЦР, 2002.

8. Фрадкина Р.Н. Говорящие руки: Тематический словарь жестового языка глухих России. // М.: Изд-во «Сопричастность» ВОИ, 2001. 598 с.

9. Селегей В.П. Электронные словари и компьютерная лексикография // AINEWS. Новости искусственного интеллекта. 2001. №1 (49). Электронный документ: http://www.lingvoda.ru/transforum/articles/pdf/selegey_a1.pdf.

10. Шиф Ж.И. Усвоение языка и развитие мышления у глухих детей. М., 1968.

11. Хахалин Г.К., Воскресенский А.Л. Контекстное фрагментирование в лингвистическом анализе // Десятая национальная конференция по искусственному интеллекту с международным участием КИИ-2006 (25–28 сентября 2006 г., Обнинск): Труды конференции. В 3 т. Т. 2. М.: Физматлит, 2006. С. 479–488.

12. Воскресенский А.Л., Хахалин Г.К. Средства семантического поиска // Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции «Диалог-2006» (Бекасово, 31 мая–4 июня 2006 г.). М.: Изд-во РГГУ, 2006. С. 100–104.

13. Жигалов В.А., Жигалов Д.В., Жуков А.А., Кононенко И.С., Соколова Е.Г., Толдова С.Ю. Система ALEX как средство для многоцелевой автоматизированной обработки текстов // Компьютерная

лингвистика и интеллектуальные технологии: Труды Международного семинара Диалог'2002. Т. 2: Прикладные проблемы. М.: Наука, 2002. С. 192–208.

14. Воскресенский А.Л., Хахалин Г.К. Кластерный анализ контекста // Международная конференция «Математическая теория систем» МТС-2009 (26–30 января 2009 г., Москва, Россия): Труды конференции. М.: ИСА РАН, 2009. С. 102–106.

15. Леонтьева Н.Н. Автоматическое понимание текстов. Системы, модели, ресурсы. М.: Издательский центр «Академия», 2006. 304 с.

16. Artificial General Intelligence. /B. Goertzel, C. Pennachin (eds). Springer, 2007.

17. Воскресенский А.Л., Воскресенский В.А., Флюр-Семенова В. Интернет-версия Каталога российских коллективов и разработок в области автоматизированной обработки речи и текстов на естественном языке // Труды Международного семинара «Диалог-2001». Т. 2. Аксаково, 2001.

18. Arapov M., Voskresensky A., Semenova V. How to compare and evaluate OCR systems? (Our approach) // Proceedings of ELSNET GO EAST and IMACS Workshop on Integration of Language and Speech. 1995, Moscow, Russia, P. 5–10.

Воскресенский А.Л.
avosj@yandex.ru

Хахалин Г.К.
gkhakhalin@yandex.ru