

Об одном подходе к выявлению инвариантных структурных единиц знаменного распева

И.В.Бахмутова, В.Д.Гусев, Т.Н.Титкова

г. Новосибирск, Институт математики им. С.Л.Соболева (ИМ СО РАН)

Bakh@math.nsc.ru, gusev@math.nsc.ru, titkova@math.nsc.ru

Аннотация. В основе изучения любой языковой системы лежит выделение структурных единиц разных иерархических уровней. В качестве языковой системы рассматриваются древнерусские музыкальные тексты, представленные в знаменной нотации. Проблема их перевода в современную нотолинейную форму носит дешифровочный характер и до сих пор остаётся актуальной.

Предложена новая система структурных единиц знаменного распева, названных нами внутригласовыми инвариантами (ВИ). На основе двознаменных певческих книг, являющихся своего рода билингвами знаменного распева, сформировано множество ВИ в виде цепочек однозначно интерпретируемых знамен. Исследована устойчивость ВИ к варьированию исходной выборки, соотнесённость их с традиционно выделяемыми уровнями иерархии знаменного распева, покрываемость ими текстов песнопений. Сделан вывод о целесообразности использования словарей ВИ в любых методиках дешифровки знаменной нотации, особенно беспометной (общий случай).

Ключевые слова: языковые системы, структурные единицы, знаменные песнопения, двознаменники, внутригласовые инварианты, проблема дешифровки.

1 Введение

Проблема перевода древнерусских церковных песнопений XII–XVII в.в., представленных в виде последовательности «знамен»¹, в современную нотолинейную форму носит дешифровочный характер и в общем случае остаётся нерешённой. Немногочисленные известные примеры дешифровок в основном относятся к текстам позднего периода, которые снабжены специальными пометами,² уточняющими нотолинейную интерпретацию знамен. Основные трудности дешифровки связаны с многозначностью соответствия «знамя – нота»: отдельные знамена могут иметь до 10 различных интерпретаций, отличающихся интервально-ритмической структурой, а также множество вариантов звуковысотной привязки. Беспометные рукописи XVI века и более раннего периода практически нечитаемы и к ним в полной мере применим термин академика Д.С.Лихачёва: «молчащая культура».

Общеязыковым приёмом, используемым для устранения неоднозначности интерпретации структурных единиц, является привлечение *контекста*. Применительно к нашему случаю это означает переход от отдельных знамен к более крупным структурным единицам – *попевкам, лицам, фитам*, с той или иной степенью полноты изученным и систематизированным в певческих азбуках, кокизниках, фитниках и различных теоретических руководствах (см. для обзора монографию М.В.Бражникова [1]). Однако даже на этом уровне неоднозначность не снимается полностью (требуется дополнительная информация о гласовой принадлежности)³

¹ Знамена (или крюки) – специальные знаки, служащие для передачи музыкальных звуков. Интерпретируются нотными цепочками переменной длины (обычно от одного до пяти нотных знаков).

² Различают *степенные* и *указательные пометы*. Первые осуществляют звуковысотную привязку распева знамен, вторые сигнализируют о возможных ритмических и высотных изменениях в распеве знамени, а также о характере исполнения (быстро, медленно, с акцентом на том или ином звуке и т.п.).

³ В древнегреческой системе осмогласия, лежащей в истоках знаменного распева, понятие *гласа* ассоциировалось с ладом (ионийский, дорийский и т.д. – всего 8 ладов). В русском осмогласии

структурной единицы, её позиции в тексте (начальная, срединная, конечная), о возможных проявлениях вариативности и т.д.).

Имеются и дополнительные обстоятельства, затрудняющие реализацию дешифровочного процесса с опорой на традиционные структурные единицы, основными из которых являются *попевки*. Укажем некоторые из этих трудностей:

– для построения формального алгоритма выделения из текста структурных единиц знаменного распева явно недостаточно тех описаний их свойств, которые содержатся в теоретических руководствах. Алгоритм, реализованный нами на основе общезыковых принципов [2], даёт хорошие результаты на массовых попевках гласа, но допускает ошибки при выделении коротких низкочастотных попевок;

– имеющиеся подборки структурных единиц представлены в форме, малоприспособленной для дешифровки *беспометной* нотации. Например, в наиболее крупной из известных подборок попевок, собранной В.М.Металловым [3], попевки представлены только в нотолинейной форме. Начертания лиц и фит знаменного распева⁴, систематизированные в работе М.В.Бражникова [4], содержат пометы (преимущественно степенные), но при этом только для половины этих структурных единиц дана нотолинейная интерпретация;

– вышеуказанные и другие (менее крупные) подборки структурных единиц не обладают полнотой, требуемой для целей дешифровки. Например, проведенный нами анализ двознаменников – певческих книг, представленных параллельно в знаменной и нотолинейной форме, показывает, что покрытие нотных текстов попевками из подборки Металлова составляет в среднем (по разным гласам) порядка 25 – 30%. Существенную роль в столь невысоком показателе играет игнорирование фактора вариативности.

В данной работе предлагается *система новых структурных единиц*, названных нами *внутригласовыми инвариантами*, которая устраняет некоторые из перечисленных выше трудностей и может быть использована для дешифровки *беспометной* нотации. В её основе лежит выделение цепочек знамен, являющихся наиболее устойчивыми (наименее вариативными) элементами знаменной нотации в каждом гласе. Применительно к знаменам, составляющим цепочку, последнюю можно трактовать как контекст, достаточный для однозначной интерпретации этих знамен в данном гласе.

2 Система обозначений.

В работе использованы следующие обозначения.

Обиходный звукоряд представлен нотами: G, A, H (малой октавы), c, d, e, f, g, a, b (первой октавы), C, D (второй октавы). В пометных рукописях высота знамен обозначается степенными пометами: Г, Н, Ц, Г (или ГН), •, н, м, п, и, м, п, и. Особенности нотолинейной интерпретации знамен поясняются с помощью системы указательных помет: Т (или Т) – тихая, Л – ломка, Б – борзая, У – ударка, К – качка (или купно), З – зевок, Р – равно. Распевы знамен с указательной пометой и без неё могут отличаться, поэтому мы считаем такие знамена разными. Более детально о семантике степенных и указательных помет см. [1, а о практике их использования – [5] и [6].

Длительности звуков обозначаются следующим образом: ○ – 1 (целая), ◐ – 2 (половинная), ◑ – 4 (четвертная), ◒ – 8 (восьмая). Для обозначения высоты и длительности звука используем комбинацию буквы и цифры (например, Н4 – это четвертная нота «си» малой октавы). При формальном сопоставлении нотолинейных цепочек (см. далее раздел 3) такая комбинация буквы и цифры, характеризующая конкретную ноту, рассматривается как один символ. Интервалы (число ступеней между высотами соседних звуков) кодируются целым числом (1 – секунда, 2 – терция, 3 – кварта и т.д.), которое сопровождается знаком (+) для восходящего движения и (–) – для нисходящего. Например, (3+) — это скачок на кварту вверх, (2–) — на терцию вниз. При повторении звука на одной и той же высоте величина интервала условно обозначается (0+). Заметим, что при таком кодировании не указывается число тонов и полутонов, содержащееся в интервалах, однако при наличии нотолинейного текста есть возможность отличать, например, большую секунду d4e4 от малой секунды e4f4 или малую терцию e2g2 от большой f2a2. Знак (*) используется в качестве разделителя между нотолинейными интерпретациями разных знамен, а (~) – как символ эквивалентности, отделяющий знаменную цепочку от её нотолинейного представления. Например, запись

(8 гласов) понятие лада деформировалось, а средством мелодической характеристики гласа стала выступать система попевок.

⁴ Условным аналогом их можно считать идиомы в естественном языке.

$\text{H} \begin{array}{|c|} \hline \text{Ш} \\ \hline \end{array} \text{А} \text{А} \sim d4c4d2*c4e4$ означает, что цепочка из двух знамен, стоящая слева от знака (\sim), интерпретируется в тексте двознаменника, соответственно, цепочками из трёх и двух нот, отделёнными друг от друга знаком (*).

3 Основы предлагаемого подхода.

Термин «*внутригласовый инвариант*» (*ВИ*) носит условный характер. Под ним мы понимаем любую связную цепочку *беспометных знамен* произвольной длины L ($L=1,2,3\dots$), имеющую одинаковую интерпретацию во всех своих вхождениях в песнопения данного гласа. Условность понятия *внутригласовый инвариант* связана с его зависимостью от исходной (в определённом смысле «обучающей») подборки, хотя в явном виде элементы *ВИ* в ней не выделены. В разделе 4 этот вопрос обсуждается подробнее.

Наборы *ВИ* для разных гласов могут частично пересекаться. Важно подчеркнуть, что речь идёт о *повторяющихся структурах*, поскольку любая *однократно* встречающаяся в подборке цепочка знамен и любые её расширения по определению имеют единственную интерпретацию.

В качестве *обучающей подборки* используются тексты из двознаменников – своего рода «билингв» знаменного распева, где они записаны параллельно *в знаменной и нотолинейной форме*. Обычно двознаменники датируются концом XVII-го – началом XVIII-го века, число их невелико, но достаточно для наших целей.

В нашем распоряжении имелись три списка двознаменной певческой книги «Октоих» вышеуказанного периода, хранящиеся в Российской Национальной библиотеке, г. С. – Петербург (Соловецкое собрание, шифры 619/647, 618/644 и QI 188). Для проведения компьютерного анализа они были нами переведены в электронную форму. В среднем гласы «Октоихов» (шифры 619/647 и 618/644) были представлены 25–27 пометными песнопениями каждый; в QI 188 этот показатель примерно на треть ниже. Длины песнопений колеблются в диапазоне от нескольких десятков до 200 и более знамен.

Множество *ВИ* формировалось для каждого из трёх «Октоихов» отдельно и в каждом из них разбивалось на подмножества $ВИ_1, \dots, ВИ_8$, соответствующие отдельным гласам. Сравнение результатов позволяет оценить устойчивость методики и выявить характер вариативности изучаемых объектов. Заметим, что ранее, на основе тех же «Октоихов», нами были построены *электронные азбуки знаменного распева* [7], содержащие полные количественные характеристики встречаемости отдельных знамен и всевозможных их интерпретаций во всех гласах. Именно на этом этапе было замечено, что, несмотря на многообразие интерпретаций у большинства знамен, существуют знамена с одинаковой нотолинейной интерпретацией во всех песнопениях своего (одного) гласа. В нашей терминологии их можно назвать *внутригласовыми инвариантами длины 1*. Тогда же и встал вопрос о существовании цепочек знамен длины 2 и выше, обладающих этим же свойством. Ответ на него даёт данная работа.

Процедура выявления *ВИ* произвольной длины состоит из трёх этапов.

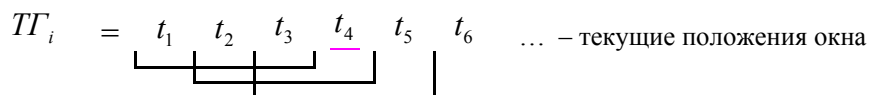
Этап 1. Устраняем все степенные и указательные пометы, стоящие у знамен⁵. Тем самым в один класс будут объединены знамена, существенно отличающиеся по своему распеву и звуковысотной привязке. Например, знамя «крюк светлый» ($\text{Ш} \begin{array}{|c|} \hline \text{Ш} \\ \hline \end{array}$), оно же с указательной пометой «ломка» ($\text{А} \begin{array}{|c|} \hline \text{Ш} \\ \hline \end{array}$), оно же с «отсекой» ($\text{Ш} \begin{array}{|c|} \hline \text{Ш} \\ \hline \end{array}$) или «ударкой» ($\text{Ш} \begin{array}{|c|} \hline \text{Ш} \\ \hline \end{array}$) при наличии у этих знамен любой степенной пометы (здесь они не поставлены) или её отсутствия попадут в один класс. После формирования множества *ВИ* мы будем в состоянии оценить, в какой степени привлечение контекста будет способствовать восстановлению утраченной информации.

Объединяем песнопения i -го гласа в один текст $ТГ_i = T_{i1} @ T_{i2} @ \dots T_{ik(i)}$, где $k(i)$ – число песнопений, представляющее i -й глас, T_{il} – песнопение с номером l ($1 \leq l \leq k(i)$), записанное в виде цепочки знамен, «@» – символ – разделитель, не принадлежащий алфавиту знамен. Каждое знамя из текста T_{il} сопровождается *нотолинейной интерпретацией*, заимствованной


⁵ При наличии нотолинейного текста они являются избыточными, поскольку лишь уточняют распевы многозначных знамен, выполняя тем самым функции того минимального контекста, который необходим для однозначной интерпретации знамени в конкретной позиции и который мы пытаемся определить.


из двознаменника. Интерпретация может меняться в зависимости от позиции знамени в песнопении.

Этап 2. Для каждого текста $ТГ_i$ ($1 \leq i \leq 8$) подсчитываем внутригласовую встречаемость всевозможных цепочек длины L , не содержащих разделитель, ($L=1,2,3,\dots,L_{i\max}$, где $L_{i\max}$ – длина максимальной повторяющейся цепочки в $ТГ_i$). Процесс ведётся итеративно по L . На L -ой итерации проходим вдоль текста скользящим окном, охватывающим L знамен, сдвигая его каждый раз на 1 символ (см. пример для $L=3$, показанный на рисунке ниже)



Сравнение и подсчёт выделяемых окном цепочек (L -грамм) осуществляется с помощью эффективной процедуры рекуррентного хеширования [8] за один просмотр текста без непосредственного сопоставления каждой L -граммы с каждой. По завершению L -ой итерации получаем полный спектр L -грамм, представленных в песнопениях i -го гласа, с указанием частоты встречаемости каждой цепочки и списка её возможных интерпретаций.

Этап 3. Для каждого гласа выделяем из списка найденных L -грамм ($L=1,2,3,\dots$) лишь те, что имеют частоту встречаемости $F \geq 2$ и единственную интерпретацию. Они и образуют *списки внутригласовых инвариантов* $ВИ_i$ ($1 \leq i \leq 8$) длины 1, 2 и т.д. Например, биграмма  встретила 15 раз в песнопениях первого гласа двознаменника 619/647 и во всех своих вхождениях имела одинаковую интерпретацию: g2*e4f4d4. Она попадает в список $ВИ_1$.

Другая же биграмма из этого гласа –  – встретила 21 раз, при этом в 19 случаях она имела интерпретацию d4c4*d4e4f2, а в двух – f4e4*d4e4f2. В силу наличия двух различных интерпретаций она не является $ВИ$.

4 Экспериментальные результаты.

Словари $ВИ_1$, сформированные по вышеописанной методике, будем иллюстрировать на материале «Октоиха» (шифр 619/647). Основные выводы таковы.

1). Число $ВИ$ разной длины во всех гласах довольно велико. В таблице 1 приведена информация о количестве $ВИ$ с частотой встречаемости $F \geq 3$, выявленных при разных L в каждом из 8 гласов.

Таблица 1. Число различных внутригласовых инвариантов разной длины в «Октоихе» (шифр 619/647) ($F \geq 3$)

Длина L	Г Л А С Ы							
	1	2	3	4	5	6	7	8
1	10	12	9	12	11	8	13	9
2	66	49	49	77	55	66	82	73
3	79*	58*	67*	96*	81*	73*	109*	95*
4	72	52	67*	82	81*	52	98	84
5	57	32	48	77	65	41	76	68
6	43	18	31	50	44	27	44	46
7	25	3	15	27	28	15	27	30
8	12	–	6	14	16	5	12	18
9	5	–	3	6	6	1	3	11
10	3	–	2	4	2	–	–	8
11	1	–	1	3	1	–	–	6

Примечание: В гласах 4 и 8 инвариантные цепочки встречаются в малом количестве и при $11 < L \leq 16$ (данные для них здесь не представлены).

Длины $ВИ$ меняются от 1 (это отдельные знамена, однозначно интерпретируемые в гласе) до 16 (это, как правило, устойчивые комбинации из двух попевок – фрагментов текста). С увеличением L число $ВИ$ растёт, при $L=3 \div 4$ достигает максимума (соответствующие значения

отмечены звёздочками), а затем монотонно падает. Такая зависимость числа *ВИ* от *L* объясняется взаимодействием двух факторов. *Первый* из них связан с увеличением многообразия *L*-грамм при переходе от *L=1* к *L=2*, от *L=2* к *L=3* и т.д. Этот фактор способствует увеличению числа *ВИ* с ростом *L* и наиболее ярко проявляет себя при переходе от *L=1* к *L=2*. Например, в гласе 1 было всего 10 знамен (*L=1*), отнесённых к *ВИ*. При *L=2* выявлены уже 66 *ВИ*⁶, при этом лишь 23 из них являются «наследственными», т.е. полученными в результате лево – или правосторонних расширений инвариантов длины 1. Остальные 43 *ВИ* длины 2 образованы сочетаниями знамен, каждое из которых в отдельности не является инвариантом, но образует таковой в комбинации друг с другом.

Второй фактор способствует уменьшению числа *ВИ* с ростом *L* и имеет простое объяснение. Частоты встречаемости *L*-грамм в тексте с увеличением *L* убывают. Поскольку в определении *ВИ* фигурируют хотя бы дважды повторяющиеся цепочки знамен, снижение частоты встречаемости цепочки до *F=1* по мере её расширения автоматически выводит *L*-грамму из числа возможных инвариантов. Анализ зависимости числа *ВИ* от *L* показывает, что для малых *L* превалирует первый фактор и число *ВИ* растёт, при больших *L* – второй и число *ВИ* падает.











2). При *L=1* в качестве *ВИ* выделяется незначительная часть (примерно одна десятая) элементов алфавита, реально встречающихся в текстах двознаменников. В таблице 2 для иллюстрации приведены *ВИ* длины 1 и 2, наиболее часто встречающиеся в первом и пятом гласах двознаменника.

Таблица 2. Примеры наиболее частых *ВИ* длины *L=1* и *L=2* для 1 и 5 гласов «Октоиха» (шифр 619/647).

ГЛАС 1				ГЛАС 5			
<i>L</i>	<i>ВИ</i>	<i>F</i>	Интерпр-я	<i>L</i>	<i>ВИ</i>	<i>F</i>	Интерпр-я
1		22	A1	1		26	c1
1		22	e4f4g2	1		14	e4f4g2
1		11	c2d2	1		9	g1
1		11	e2f4e4	1		9	f1
1		9	a8g8f4g4f4	1		8	a4g4f4a4
1		8	d4e4f2	1		6	d2c2
1		8	f4g4a2	1		6	g4a4f4
1		5	e1	1		5	d4e4f2e2d2
2		27	e2*f4e4	2		44	d1*c1
2		25	f4e4*d2	2		32	f4g4*a2.g4
2		15	g2*e4f4d4	2		29	f4e4*d1
2		13	e1*d1	2		29	d2*c4f2e4
2		13	e2f4g4*e2	2		25	d1*c1
2		12	A1*G1	2		16	f4g4a4f4*g1
2		11	c4d4e4d4*e1	2		15	g2a4b4*a4g4
2		10	g2*e4f4e4d4	2		14	g2.f4*g2











Анализ таблицы 2 показывает, что среди *ВИ* встречаются не только низкочастотные *L*-граммы, но и относительно высокочастотные (см., например, и (гл. 5), (гл. 1) и др.). Пересечений между множествами *ВИ* одинаковой длины немного, даже для гласов близких в мелодическом отношении (такими считаются гласы 1 и 5, 2 и 6, 3 и 7, 4 и 8). При *L=1* в пересечение попадает знамя «ключ» (), а при *L=2* – цепочки и .










⁶ Напомним, что в таблице 1 не представлены данные о *ВИ* с частотой *F=2*. Их включение увеличивает количество цепочек, формально относимых к *ВИ*.





По таким цепочкам можно отслеживать *гласоспецифичность* знаменного распева. Интересным представляется и то, что частоты биграмм, возглавляющих топ-списки *ВИ*, в среднем выше, чем частоты однограммных *ВИ*. Это подтверждает тот факт, что при переходе от $L=1$ к $L=2$ число *ВИ* растёт, в основном, не за счёт «наследственных» L -грамм, являющихся расширениями *ВИ* длины 1 (см., например, пары  и  или  и  (глас 1);  и  (глас 5)), а в результате комбинирования однограмм, не являющихся *ВИ* (см., например, ,  и др. (глас 1) или ,  и др. (глас 5)). С увеличением L ситуация меняется: при переходе от L к $L+1$ начинают преобладать «наследственные» L -граммы, т.е. имеет место вложенность одних *ВИ* в другие.

3). Важным является вопрос, что же представляют собой *ВИ* в функциональном отношении, как они соотносятся с традиционно выделяемыми структурными единицами знаменного распева, в первую очередь, с попевками. Одна из классификаций попевок была предложена А.Н.Кручининой [9]. Согласно этой классификации все попевки разбиваются на 24 семейства в соответствии с тремя знаменами, завершающими попевку (так называемый «каданс»), реализующий ритмический останов мелодии и подчёркивающий опорные ступени гласа). Кадансовая структура считается наиболее устойчивым элементом попевки. Естественно было ожидать, что короткие *ВИ* ($L=1÷3$) будут являться элементами каданса, поскольку предшествующая кадансу часть попевки – «подвод» – существенно более вариативна.





Анализ таблицы 2 и других данных, не приводимых здесь ради экономии места, показывает, что это предположение оправдывается лишь частично, в первую очередь для наиболее часто встречающихся коротких *ВИ*. Среди таковых в таблице 2 отметим знамена





 и  (гл. 1), ,  (гл. 5), интерпретируемые целой нотой (ритмический останов). Среди *ВИ* длины 2 кадансовыми являются цепочки: , ,  в гласе 1 и , ,  в гласе 5.

Однако некоторые короткие *ВИ* не являются элементами кадансовой структуры. Например, самый частый *ВИ* длины 2 в первом гласе –  ($F=27$) – выполняет роль универсального подвода в попевках «хамила»:   ($F=4$) и «кокиза»:   ($F=14$)⁷. Это *ВИ* минимальной длины, встретившийся в указанных попевках. С ним в тех же попевках «зацеплен» (пересекается) другой *ВИ* длины 2 – , имеющий чуть меньшую частоту ($F=25$). Аналогично, в пятом гласе *ВИ* длины 2 –  ($F=15$) играет роль подвода в попевке «долинка»:   ($F=8$). Число подобных случаев увеличивается с уменьшением частоты *ВИ*. Более того, иногда *ВИ* длины 2 (или больше) образуется на стыке двух попевок, объединяя последнее (кадансовое) знамя первой попевки и начальное (переходное) второй попевки. В качестве примера можно привести два *ВИ* из пятого гласа:


 ~ g1*f4g4a2 ($F=5$) и  ~ a1*g4a4 ($F=5$). Подобные случаи легко трактуются: вслед за кадансом следует стандартный переход к другой попевке. Явления такого рода отражены в так называемых «составных» знаменах типа  (статья с переводкой), в котором кадансовое знамя () интерпретируется половинной, а не целой нотой.


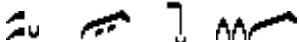
ВИ средней длины ($4 \leq L \leq 8$) представлены преимущественно либо законченными попевками, либо достаточно крупными их фрагментами, зацепленными друг с другом. Примером первых являются цепочки:

    ~ a2*g4f4 * g2 * f2 * e1 – «кокиза», гл.1 ($F=11$);



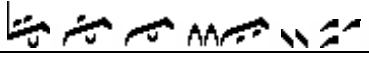

    ~ e4f4 * g2 * e4f4e4d4 * c4d4c4H4 * A1 – «кулизма», гл.1 ($F=10$);

⁷ Здесь и далее названия попевок даются в соответствии с классификацией А.Н.Кручининой [9].

 ~ g4f4 *g2 *f4e4 *d2 *c4f2e4 *d1 *c1 – «долинка», гл.5 (F=13) и др.
Примером зацепленных фрагментов (незавершенные попевки) могут служить:

 ~ a4g4 * g4e4f4e4 * f4g4 * f4e4 (F=10, гл.5)
 ~ g4e4f4e4 * f4g4 * f4e4 d1 (F=12, гл.5)

Длинные *ВИ* ($L \geq 9$) обычно соответствуют устойчивым комбинациям двух попевок или их незавершенным сдвиговым вариантам. Приведем примеры таких комбинаций:

 «прикладная» +  «грунка» ~
e2 *e4d4 *e1 *d4c4H2 * e2 *f4e4 *d4e4f2 *e1 *d1 (F=3, гл.4);
 «сложитие» +  «путик» ~
g2f2 *f4e4 *d4c4 *d4e4f4. *e8d4 *e1 * e2d2 *c2 *d4e4 *d1 (F=3, гл.8);

Проводя аналогию с естественным языком, *ВИ* малой длины можно соотнести с морфемами, *ВИ* средней длины (попевки) – со словами, а супердлинные *ВИ* (комбинации попевок) – с устойчивыми словосочетаниями. Используя эту аналогию, можно сказать, что выделяемые нами структурные единицы затрагивают три нижних уровня той специфической языковой системы, которая называется знаменным распевом.

5 Оценки покрываемости песнопений внутригласовыми инвариантами.

Из определения *ВИ* не следует, что полная их совокупность будет покрывать весь исходный текст. Решалась задача выделения наиболее информативных (консервативных) фрагментов, а не задача сегментации текста, т.е. разбиения его на части в соответствии с некоторым критерием. Поэтому представляет интерес оценить, какая доля исходного (или контрольного) текста покрывается выделенными *ВИ*. Коэффициент покрытия конкретного песнопения подсчитывается следующим образом. Выявляем в песнопении все цепочки знамен, соответствующие вхождению внутригласовых инвариантов из сформированного словаря *ВИ*. Подсчитываем суммарное число позиций n , занимаемых этими цепочками в песнопении. Каждая позиция учитывается лишь *один раз*. Тем самым нивелируется эффект вложенности или зацепленности (перекрываемости) отдельных *ВИ*. Коэффициент покрытия k определяется как отношение n/N , где N – число знамен в песнопении (его длина).

Коэффициенты покрытия отдельных песнопений в каждом гласе меняются в широких пределах: в первом гласе от 0.21 до 0.85, во втором – от 0.30 до 0.73, в третьем – от 0.29 до 0.83 и т.д. Средние значения коэффициента покрытия \bar{k} по каждому гласу двознаменника (шифр 619/647) приведены ниже (порог отбора *ВИ* по частоте равен 3):

Глас	1	2	3	4	5	6	7	8
\bar{k}	0.636	0.487	0.615	0.676	0.625	0.512	0.798	0.66

Нетрудно видеть, что в среднем коэффициенты покрытия песнопений каждого гласа внутригласовыми инвариантами *достаточно высоки* – порядка 60% и выше⁸. Несколько выпадают из общей картины два близких (параллельных) гласа – второй и шестой – с более низкими значениями \bar{k} , а также глас 7 – с более высоким коэффициентом покрытия. Значительное различие значений k для отдельных песнопений каждого гласа обусловлено,

⁸ Напомним, что аналогичные показатели для покрытия нотолинейной составляющей двознаменников попевками из подборки В.М.Металлова были порядка 30% и ниже.

скорее всего, жанровой спецификой⁹. Детальный разбор отмеченных особенностей требует специального рассмотрения.

Приведённые выше показатели относились к случаю, когда покрываемость песнопений оценивалась на том же самом «обучающем» материале, который использовался при построении множества *ВИ*. Для получения более реалистичной оценки был проведён эксперимент в режиме скользящего контроля. Из обучающей подборки песнопений гласа удалялось одно песнопение («контрольное»), а по оставшимся строилось множество *ВИ*. Подсчитывался коэффициент покрытия контрольного песнопения выделенными *ВИ*. Затем песнопение возвращалось в обучающую подборку, а в качестве «контрольного» рассматривалось следующее песнопение, не анализировавшееся ранее. Процесс повторялся до исчерпания всех песнопений гласа. Средние значения коэффициента покрытия \bar{k} по каждому гласу двознаменника (шифр 619/647), полученные в режиме скользящего контроля, приведены ниже:

Глас	1	2	3	4	5	6	7	8
\bar{k}	0.555	0.384	0.511	0.506	0.566	0.378	0.695	0.538

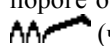

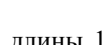

Можно видеть, что в режиме скользящего контроля наблюдается вполне ощутимое снижение показателей покрываемости по сравнению с предыдущим случаем.


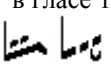
Ещё один эксперимент по оцениванию покрываемости песнопений внутригласовыми инвариантами был проведён с привлечением «Октоиха» (618/644), песнопения которого рассматривались в качестве «контрольного» материала, а обучение, т.е. построение множества *ВИ*, проводилось на основе «Октоиха» (619/647)¹⁰. Отметим, что несмотря на близость обоих двознаменников по составу песнопений, в них отмечаются довольно существенные различия на знаменном и нотолинейном уровне (сравнительный анализ был проведён в [7]). Приведённые ниже средние значения коэффициентов покрытия песнопений из «Октоиха» (618/644) внутригласовыми инвариантами, построенными на основе «Октоиха» (619/647), подтверждают выводы из [7]:

Глас	1	2	3	4	5	6	7	8
\bar{k}	0.492	0.326	0.410	0.372	0.432	0.218	0.542	0.459

Эти результаты показывают, что имеет место дальнейшее снижение показателей покрываемости даже по сравнению с вариантом скользящего контроля, т.е. степень различия между двумя списками «Октоихов» довольно велика. Различия могут быть обусловлены как эволюционными факторами («Октоихи» разнесены во времени), так и диалектными, возможно привнесёнными переписчиками – приверженцами разных певческих школ.

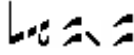



6 Обсуждение результатов.

6.1. С первого взгляда может показаться, что методика выделения *ВИ* очень чувствительна к всевозможным проявлениям вариативности знаменного распева. Действительно, если сравнить, например, множества *ВИ* длины 1 для первых гласов «Октоихов» (619/647 и 618/644) при пороге отбора $F=3$, то общими для этих множеств будут лишь 3 знамени: «стрела громная» –  (частоты встречаемости в «Октоихах», соответственно, 22 и 25), «стрела громомрачная» –  (частоты 8 и 18) и «дербица» –  (частоты 3 и 3). Наиболее высокочастотный *ВИ* длины 1 из «Октоиха» (618/644) – «переводка с подверткой» –  ~ e4f4d4, $F=24$ – в этом списке отсутствует, поскольку в «Октоихе» (619/647) это знамя не является инвариантом в гласе 1. Оно имеет две разные интерпретации: e4f4d4 ($F=18$) и d4f4d4 ($F=3$).

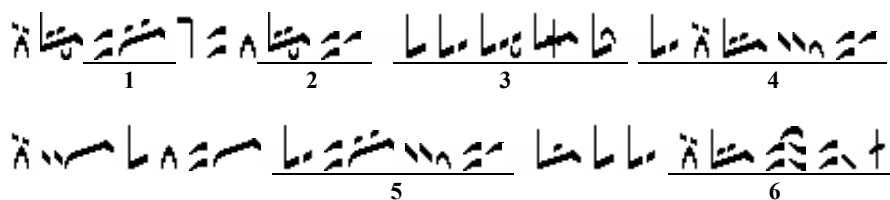
Проследим за всеми лево- и правосторонними расширениями знамени  в гласе 1 «Октоиха» (619/647). При левостороннем расширении мы фиксируем *ВИ* длины 2:  ~g2 e4f4d4




⁹ «Октоих» – многожанровая книга песнопений.



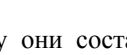
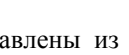
¹⁰ *ВИ*, полученный на основе «Октоиха» (619/647), может не являться таковым для «Октоиха» (618/644). В этом случае при вычислении покрытия учитывались лишь те вхождения *ВИ* в тексты контрольного «Октоиха», которые имели ту же самую, что и в (619/647), нотолинейную интерпретацию.

($F=15$), а при правостороннем – два *ВИ* длины 3:  ~ e4f4d4*e1*d1 ($F=9$) и  с той же интерпретацией ($F=6$). Эти *ВИ* возникли в результате комбинации знамен, по отдельности не являющихся *ВИ*. Заметим, что аналогичные цепочки представлены среди *ВИ* длины 3 в «Октоихе» (618/644) с частотами ($F=13$) и ($F=6$) соответственно, но в этом «Октоихе» они носят наследственный характер (см. раздел 4). Таким образом, отсутствие знамени  среди *ВИ* длины 1 в «Октоихе» (619/647) не явилось существенным препятствием к однозначной идентификации кадансовой структуры попевки «колесо»  в обоих «Октоихах». Описанный механизм обеспечивает достаточный «запас прочности» любой методике дешифровки, основанной на использовании *ВИ*.

6.2. Внутригласовые инварианты можно трактовать как фрагменты песнопений, наилучшим образом сохранившиеся в ходе эволюции знаменного распева. Ниже для иллюстрации приведена последовательность знамен из «антифона» 7-го гласа, где эти фрагменты (режим скользящего контроля)¹¹ выделены подчеркиванием:



Коэффициент покрытия достаточно высокий: $k \sim 0.63$. Результаты выглядят довольно логично: в качестве *ВИ* выделены законченные попелочные структуры (см. фрагменты №№3,4,5,6, при этом 4-ый и 5-ый совпадают на нотолинейном уровне). Фрагменты №№1 и 2 – элементы кадансовых структур, соответственно,  и . Эти структуры интерпретируются однозначно, но не включены в число *ВИ*, поскольку встречаются в песнопениях седьмого гласа по разу и не проходят порог отбора ($F_{пор.}=3$). По той же причине не выявлена короткая кадансовая структура , встретившаяся в 7-ом гласе один раз.

Цепочки  и  также остались непокрытыми *ВИ*, что вполне объяснимо, поскольку они составлены из высокочастотных знамен (, ) с широким диапазоном звуковысотных привязок внутри гласа. Обычно они встречаются на стыках структурных единиц и отличаются повышенной вариативностью.

Анализ покрытия конкретного песнопения внутригласовыми инвариантами даёт объективную оценку того, какие знамена в песнопении следовало бы снабжать пометами с целью уточнения их интерпретации. Заметим в связи с этим, что расстановка помет в одних и тех же песнопениях из разных «Октоихов» существенно различается, что свидетельствует об отсутствии единодушия в этом вопросе.

6.3. Описанная в работе система *ВИ* представляет непосредственный интерес для решения задачи дешифровки беспометных песнопений. Достаточно высокие средние значения коэффициентов покрытия по разным гласам гарантируют наличие множества опорных (однозначно интерпретируемых) фрагментов внутри песнопения. Напомним, что приведённые выше результаты по коэффициентам покрытия были получены при пороге отбора *ВИ* по частоте, равном 3. Формально могут быть использованы и *ВИ* с частотой 2 и даже 1 (фрагменты, однократно встречающиеся в песнопениях гласа). Это повышает показатели покрытия, но уменьшает надёжность предлагаемого решения.

Дальнейшими шагами, связанными с повышением эффективности подхода, основанного на *ВИ*, являются: расширение объёмов обучающих подборок (двознаменников), более детальный учёт жанровой специфики, качественное и количественное исследование проявлений вариативности, привлечение дополнительной информации, почерпнутой из музыкально-теоретических пособий и подборок структурных единиц знаменного распева.

¹¹ При наличии вложенности использовался *ВИ* максимальной длины. Разрывы в подчеркивании смежных фрагментов означают нерасширяемость соответствующих *ВИ*.

Заключение

Исследуется специфическая система структурных единиц знаменного распева, названных нами *внутригласовыми инвариантами*. В её основе лежит выделение цепочек знамен различной длины, *однозначно интерпретируемых* (в пределах исходной обучающей подборки) *в конкретном гласе*. В качестве обучающей подборки использованы разные варианты (списки) двознаменной многожанровой певческой книги «Октоих» конца 17-го – начала 18-го века.

Построены множества *ВИ* для всех восьми гласов «Октоиха» и исследована их устойчивость в экспериментах со скользящим контролем и с привлечением материала, не использовавшегося при обучении. Показано, что с помощью *ВИ* покрывается (а, следовательно, по аналогии с обучающей выборкой и интерпретируется) в среднем около 50% текста в *беспометных песнопениях*. Места вхождения *ВИ* в текст можно трактовать как своего рода «островки надёжности», характеризующиеся существенным снижением уровня неоднозначности в интерпретации знамен. Благодаря этой особенности *ВИ* могут быть использованы в любых методиках дешифровки беспометной нотации и представляют значительный интерес в плане изучения вариативности знаменного распева.

Литература

- [1] Бражников М.В. Древнерусская теория музыки, Л.: изд. «Музыка», 1972, – 422 С.
- [2] Бахмутова И.В., Гусев В.Д., Титкова Т.Н. L-граммные азбуки для дешифровки знаменных песнопений // *Сибирский журнал индустриальной математики*. – т.1, N2, с.51–66. Новосибирск 1998.
- [3] Металлов В.М. Осмогласие знаменного распева (сборник нотолинейных попевок), с.1–50, М.: –1899 г.
- [4] Бражников М.В. Лица и фиты знаменного распева, Л.: изд. «Музыка», 1984,– 302 С.
- [5] Бахмутова И.В., Гусев В.Д., Титкова Т.Н. О функциях указательных помет (на материале двознаменника XVIII века) // *Сибирский музыкальный альманах*. – Изд-во НГК, с. 81–92. Новосибирск 2002.
- [6] Бахмутова И.В., Гусев В.Д., Титкова Т.Н. Факторы, влияющие на точность нотолинейной реконструкции пометных знаменных песнопений // *Сибирский музыкальный альманах*. – Изд-во НГК, с. 51–59. Новосибирск 2004.
- [7] Бахмутова И.В., Гусев В.Д., Титкова Т.Н. Электронная азбука знаменного распева: Предварительная версия. // *«Анализ структурных закономерностей», Вычислительные системы*, вып.174, С. 29–53. Новосибирск 2005.
- [8] Гусев В.Д., Титкова Т.Н., Рекуррентное хеширование символьных цепочек. // *«Анализ последовательностей и таблиц данных», Вычислительные системы*, вып. 150, С. 94—106. Новосибирск 1994.
- [9] Кручинина А.Н. Попевка в русской музыкальной теории XVII века: Дис. ... канд. искусствовед. наук: 17.00.02 — Л., 1979.