

Выявление инвариантов и квазиинвариантов знаменного распева с помощью билингв типа "знамя-нота"¹

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 нотоплинейную форму носит дешифровочный характер и до сих пор остаётся актуальной.

В дополнение к предложенным ранее в [7] внутригласовым инвариантам (ВИ), выступающим в качестве новой системы структурных единиц знаменного распева, предлагается использовать квазиинварианты (КВИ). ВИ – это повторяющиеся цепочки разной длины, однозначно интерпретируемые в конкретном гласе. КВИ – это тоже повторяющиеся цепочки разной длины, но имеющие несколько нотоплинейных интерпретаций. Предложен критерий отбора доминирующей интерпретации.

На основе трёх двознаменников (тексты в формате «знамя – нота») построены множества ВИ и КВИ, исследованы их характеристики, получены оценки покрываемости этими цепочками беспометных контрольных песнопений. Отмечены достоинства и недостатки предлагаемых характеристик. Словари ВИ и КВИ могут быть использованы в любых методиках дешифровки беспометной нотации.

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



1 Введение

Древнерусские церковные песнопения XII–XVIII в.в. чаще всего представлены в знаменной форме записи, являвшейся ведущей в системе музыкальной письменности того времени. Знамена – специальные знаки, служащие для передачи мелодии. При переводе знамен в современную нотоплинейную форму они интерпретируются цепочками нот разной длины (от одного до пяти–шести нотных знаков). Проблема перевода песнопений из знаменной формы записи в современную нотоплинейную носит дешифровочный характер и в общем случае остаётся нерешённой. Немногочисленные известные примеры нотоплинейной реконструкции знаменных песнопений относятся к так называемым "пометным текстам", где знамена снабжены дополнительными знаками (пометами), облегчающими их интерпретацию. Различают степенные и указательные пометы. Первые уточняют высоту распева знамени, а вторые – особенности его распева, которые можно трактовать как отклонения от стандартного варианта. Система помет была разработана в XVII веке. Беспометные певческие рукописи XVI века и более раннего периода практически не читаемы.

В истоках знаменного распева лежит древнегреческое церковное пение, перенесённое на русскую почву после принятия христианства. Оно регламентировалось достаточно сложной системой осмогласия, которая подразумевала пение на 8 ладов (гласов): ионийский, дорийский и т.д. В русском осмогласии понятие лада деформировалось, а средством мелодической характеристики гласа стала выступать система попевок – основных структурных

¹ Работа выполнена при поддержке гранта РФФИ №13-07-00400 "Формирование электронных словарей инвариантов, квазиинвариантов и попевок для дешифровки знаменных песнопений"

единиц (СЕ) знаменного распева. Более детальные сведения о специфике древнерусской музыки могут быть почерпнуты из монографии М.В.Бражникова [1].

Основные проблемы дешифровки связаны с *многозначностью* соответствия «знамя – нота». Одна и та же цепочка знамен может иметь разные интерпретации в зависимости от гласа, типа СЕ, в состав которой она входит, позиции в тексте и ряда других факторов. Например, знамя "крюк светлый" ()¹⁰, распеваемое обычно одной половинной нотой, при наличии у него указательной пометы "ломка" ()¹⁰ интерпретируется в первом и пятом гласах уже двумя четвертными в восходящем движении, причём их звуковысотные привязки в этих гласах не совпадают. Наличие помет облегчает дешифровку, но не гарантирует, тем не менее, однозначного результата из-за наличия в пометных текстах значительного числа беспометных знамен, а также многочисленных разночтений и исключений в интерпретации самих помет [2,3].

Сложность дешифровки беспометной нотации по сравнению с пометной многократно возрастает. Для устранения неоднозначности в интерпретации знамен требуется привлечение *контекста*, что эквивалентно переходу от отдельных знамен к более крупным СЕ, таким как попевки, лица, фиты и др. Однако известные подборки СЕ (азбуки, кокизники, фитники) малоприспособлены для целей дешифровки, в первую очередь, из-за формы их представления. Например, в подборке В.М.Металлова [4] попевки приведены только в нотолинейной форме, а в подборке М.В.Бражникова [5] около половины лиц и фит имеют лишь знаменное представление.

Для создания словарей СЕ в требуемом формате "знамя–нота" необходимо привлекать *билингвы* – певческие книги, в которых песнопения представлены параллельно в знаменной и нотолинейной форме (так называемые двознаменники). Они начали появляться в XVII веке и количество их весьма ограничено. Используя двознаменники, авторы данной работы построили *электронную азбуку знаменного распева* [6], выгодно отличающуюся от известных авторских азбук по многим показателям, в частности, наличием информации о частоте встречаемости и звуковысотной привязке различных интерпретаций каждого знамени в каждом гласе. Анализ этой информации показал, что при всей вариативности знаменного распева имеются знамена, однозначно интерпретируемые в пределах одного гласа. Их естественно назвать внутrigласовыми инвариантами (*ВИ*). В работе [7] авторы показали, что в качестве *ВИ* могут выступать не только отдельные знамена, но и цепочки знамен длины L ($L = 2, 3, \dots$). Места вхождения таких *ВИ* в текст можно трактовать как своего рода "островки надёжности", характеризующиеся существенным снижением уровня неоднозначности в интерпретации знамен.

В данной работе исследуется возможность ослабления требования однозначной интерпретируемости цепочек, выделяемых в качестве *ВИ*. Это означает, что некоторая неопределённость в интерпретации цепочки сохраняется, но она носит характер допустимой (и к тому же регулируемой) вариативности. Цепочки такого типа будем называть *квазиинвариантами (КВИ)*. Целью работы является *формирование словарей КВИ* для каждого из 8 гласов на основе двознаменников конца XVII – начала XVIII века и их характеристика (объёмы словарей, межгласовые различия, функциональная нагрузка *КВИ* разной длины, покрываемость песнопений цепочками, представленными в словарях и т.д.). Важно отметить, что словари *КВИ*, равно как и словари *ВИ*, могут быть использованы для дешифровки не только пометной, но и *беспометной нотации*. Однако первые, в отличие от вторых, учитывают и проявления вариативности, характерные для СЕ знаменного распева.

2 Система обозначений.

В работе использованы следующие обозначения.

Обиходный звукоряд (см. рисунок ниже) 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: \dagger , \ddagger , ц , г (или гн), \bullet , н , м , п , п , м , п , п .



Особенности нотолинейной интерпретации знамен поясняются с помощью системы указательных помет: **Т** (или **—**) – тихая, **°** – ломка, **з** – борзая, **У** – ударка, **«** – качка (или купно), **з** – зевок, **Р** – равно. Знамя с указательной пометой и без неё трактуются как разные, поскольку могут иметь отличающиеся распевы. Более детально о семантике степенных и указательных помет см. [1], а о практике их использования – [2,3].

Длительности звуков обозначаются следующим образом: **о** – 1 (целая), **д** – 2 (половинная), **л** – 4 (четвертная), **в** – 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 конкретный звук, рассматривается как один символ. Интервалы (число ступеней между высотами соседних звуков) кодируются целыми числами (1 – секунда, 2 – терция, 3 – кварта и т.д.), которые сопровождаются 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 наличие нотолинейного текста дает 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 $\text{Н}^{\text{л}} \text{л}^{\text{°}} \text{ж} \sim \text{d4c4d2} * \text{c4e4}$ означает, что цепочка из двух знамен, стоящая слева от (~), интерпретируется в тексте двознаменника, соответственно, цепочками из трёх и двух нот, отделёнными друг от друга знаком (*)).




3 Выявление внутригласовых квазиинвариантов в текстах двознаменников.

Разработанный нами метод решения предполагает наличие достаточно *представительных подборок песнопений разных жанров* по каждому гласу, записанных параллельно в знаменной и нотолинейной форме. В нашем распоряжении имелись три варианта многожанровой певческой книги "Октоих", датируемые концом XVII– началом XVIII века (РНБ, г. С. –Петербург, Соловецкое собрание, шифры 619/647, 618/644 и QI 188). Каждое песнопение представлено: цепочкой знамен с приписанными степенными и указательными пометами²; нотолинейным текстом, разбитым на фрагменты, соответствующие распеву отдельных знамен³; стихотворным (старославянским) текстом, разбитым на слоги с привязкой к отдельным знаменам. Перевод песнопений в электронную форму и частично автоматизированный контроль ошибок осуществлён авторами данной работы. В среднем гласы двух первых двознаменников были представлены 25÷29 песнопениями каждый; в двознаменнике QI 188 этот показатель на треть ниже. Иллюстрация результатов проводится на материале двознаменника 619/647.

В основу выделения *ВИ* и *КВИ* положены сходные соображения:

² У части знамен пометы отсутствуют

³ Возможны "склеивания" интерпретаций отдельных знамен, требующие выработки стратегии их разделения.

(1) если мы хотим, чтобы формируемые словари СЕ, каковыми являются *ВИ* и *КВИ*, были пригодны для нотолинейной реконструкции беспометных текстов, нужно устранить все пометы (степенные и указательные) у знамен исходной подборки. Это равносильно агрегированию исходного знаменного алфавита, при котором объединяются в одну группу и обозначаются одним символом все варианты распева конкретного знамени. В пометных текстах эти варианты распева фиксируются, отличающимися друг от друга комбинациями указательных и степенных помет. В соответствии с этим в одну группу будут объединены, например,  (стрела мрачная крыжевая с оттяжкой), она же с указательной пометой "борзо"  и "тихая" , имеющие разные распевы. Формируемые на основе агрегированных данных словари *ВИ* и *КВИ* можно тогда трактовать как контексты, необходимые для восстановления утраченной информации;

(2) элементами словарей *ВИ* и *КВИ* могут быть лишь *повторяющиеся* цепочки знамен, поскольку понятие инвариантности подразумевает сохранение (с допустимыми отклонениями) какого-либо качества (в нашем случае нотолинейной интерпретации цепочки) при повторении события. Требование повторяемости исключает из рассмотрения однократно встречающиеся в подборке цепочки знамен, хотя формально они и любые их расширения имеют единственную интерпретацию. Это же требование является существенным в плане ограничения длин *ВИ* и *КВИ*, поскольку с увеличением длин цепочек частота их встречаемости в подборке (F) быстро падает. При $F = 1$ цепочка выбывает из рассмотрения;

(3) цепочка, составленная из многозначных (в общем случае) знамен, может оказаться однозначно (или "почти однозначно") интерпретируемой в конкретном гласе. Поэтому при формировании словарей *ВИ* и *КВИ* должен быть рассмотрен полный спектр цепочек произвольной длины, представленных в песнопениях гласа, частота встречаемости которых равна или превышает пороговое значение $F_{пор}$.

Исходя из приведённых соображений, алгоритм формирования словарей *КВИ* для разных гласов включает в себя следующие шаги:

Шаг 1. Устраняем все степенные и указательные пометы, которыми снабжены знамена. Объединяем песнопения i -го гласа ($i=1\div 8$) в один текст $ТГ_i = T_{i1} @ T_{i2} @ \dots T_{ik(i)}$, где $k(i)$ – число песнопений в i -ом гласе, T_{il} – песнопение с номером l ($1 \leq l \leq k(i)$), записанное в виде последовательности беспометных знамен, @ – символ-разделитель между песнопениями. Каждое знамя сопровождается нотолинейной интерпретацией, взятой из двознаменника. Одно и то же знамя, стоящее в разных позициях, может иметь отличающиеся интерпретации.

Шаг 2. Для каждого текста $ТГ_i$ ($1 \leq i \leq 8$) подсчитываем внутригласовую встречаемость всевозможных цепочек знамен длины L (L -граммы), 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, выделяющим L очередных знамен, сдвигая его каждый раз на один символ. Сравнение и подсчёт выделяемых окном L -грамм осуществляется с помощью процедуры рекуррентного хеширования [8] за один просмотр текста $ТГ_i$ без непосредственного сравнения L -грамм. По завершении L -ой итерации получаем полный спектр L -грамм, представленных в песнопениях i -го гласа, с указанием частоты встречаемости каждой цепочки и списка её всевозможных нотолинейных интерпретаций (тоже с частотами).

Шаг 3. Для каждого гласа выделяем из списка найденных L -грамм ($L = 1, 2, 3, \dots$) лишь те, что имеют частоту встречаемости $F \geq F_{пор} > 1$. В описываемой версии словаря *КВИ* значение $F_{пор}$ выбрано равным 3. Каждая из отобранных L -грамм может иметь разное число интерпретаций. Если интерпретация единственная, цепочка является *внутригласовым инвариантом* относительно заданной подборки $ТГ_i$. Если интерпретаций несколько, анализируем их частоты. При этом возможны разные стратегии формирования словаря *КВИ*. В первом приближении будем отбирать в этот словарь (и, соответственно, называть *КВИ*) цепочки, среди интерпретаций которых имеется *доминирующая по частоте*. Доминирующей

будем считать интерпретацию, частота встречаемости которой $F_{\text{доп}}$ превышает суммарную частоту встречаемости всех других интерпретаций анализируемой цепочки. Это означает, что выполняется соотношение: $F_{\text{доп}}/F > 1/2$, где F – частота встречаемости цепочки. В данной работе мы использовали *более жёсткое правило* доминирования, основываясь на том, что нижний порог отбора L – грамм по частоте у нас равен 3 и таких цепочек довольно много. Доминирующая интерпретация для них должна иметь частоту 2, отсюда $F_{\text{доп}}/F = 2/3$. Задав критерий отбора в виде $F_{\text{доп}}/F \geq 2/3$, для любых $F \geq 3$, мы достигаем некоторого компромисса между покрываемостью текста цепочками из словаря *КВИ* и количеством ошибок дешифровки, неизбежных при использовании *КВИ*.

4 Экспериментальные результаты.

1). По объёму словаря *КВИ* для разных гласов несколько уступают словарям *ВИ*. Это различие становится особенно заметным с увеличением L . Динамика изменения числа *КВИ* с ростом L в каждом гласе сходна с аналогичной характеристикой для *ВИ*: нарастание при малых L , максимум при $L = 3 \div 4$, убывание при $L \geq 4$ (см. обсуждение в [7]). Для $L \approx 7 \div 8$ число *КВИ* уже существенно меньше числа *ВИ*. По-видимому, это связано с тем, что *КВИ* является "менее устойчивым" объектом, чем *ВИ*, и раньше (т.е. при меньших L) выходит за пределы минимально допустимой частоты встречаемости.

2). Для иллюстрации различий между словарями *ВИ* и *КВИ* приведём две таблицы: в первой из них указаны *ВИ* и *КВИ* длины 1 для первого гласа двознаменника 619/647. Это отдельные знамена, однозначно (*ВИ*) или "почти однозначно" (*КВИ*) интерпретируемые в этом гласе. Во второй таблице представлены наиболее часто встречающиеся *ВИ* и *КВИ* для значений $L = 2 \div 4$ (глас 1). Доминирующие интерпретации для *КВИ* подчёркнуты.

Таблица 1. Фрагменты словарей *ВИ* и *КВИ* для $L = 1$
(L – длина, F – частота встречаемости в гласе)

<i>ВИ</i>	F	Интерпретация	<i>КВИ</i>	F	Интерпретация
	22	A1		230	<u>e2 -164</u> , c2 -9, d2 -31, e4 -5 и т.д.
	22	e4f4g2		78	<u>d2 -63</u> , A2 -6, e2 -3, d4 -2 и т.д.
	11	c2d2		59	<u>e1 -55</u> , d1 -2, g1 -1, a1 -1
	11	e2f4e4		36	<u>d4e4f2 -30</u> , c4d4e2 -2, H4c4d2 -4,
	9	a8g8f4g4f4		21	<u>e4f4d4 -18</u> , d4f4d4 -3
	8	d4e4f2		21	<u>e2f4g4 -15</u> , d2e4f4 -6
	8	f4g4a2		20	<u>e2 -15</u> , g2 -2, d2 -3
	5	e1		20	<u>c4d4e4d4 -14</u> , f4g4a4g4 -5, e4f4g4f4 -1
	4	e2f4g4		15	<u>f1 -13</u> , e1 -2
	3	e4f4g4a4		12	H4H4A2 -4, <u>d4e4d2 -8</u>

Таблица 2. Наиболее часто встречающиеся *ВИ* и *КВИ* для значений $L = 2 \div 4$
(L – длина, F – частота встречаемости в гласе)

L	<i>ВИ</i>	F	Интерпр-я	<i>КВИ</i>	F	Интерпретация
2		27	e2*f4e4		86	<u>e2*e2 -62</u> , d2*d2 -10, e4*e4 -1
2		25	f4e4*d2		36	<u>a2*g4f4 -25</u> , g2*f4e4 -9 и т.д.
2		15	g2*e4f4d4		34	<u>d2 e2 -32</u> , e2*f2 -2
2		13	e1*d1		22	<u>d4c4*d4e4f2 -20</u> , f4e4*d4e4f2 -2
2		13	e2f4g4*e2		21	<u>e4d4*d4H4c4H4 -17</u> , e4d4*d8c8H4c4H4 -3 и т.д.

отслеживание взаимосвязей между цепочками *ВИ* и *КВИ* разной длины может способствовать уточнению границ попевок – основных СЕ знаменного распева.







5 Оценка покрываемости песнопений цепочками из словарей *ВИ* и *КВИ*

Из определения *ВИ* и *КВИ* не следует, что полная их совокупность будет покрывать весь исходный текст (введение порога отбора по частоте, равного трём, уже не гарантирует этого). Поэтому представляет интерес оценить степень покрываемости выделенными *ВИ* и *КВИ* как исходных текстов, так и контрольных, не использовавшихся при формировании словарей *ВИ* и *КВИ*. Оценки покрываемости с помощью *ВИ* приведены нами в [7]. Здесь же мы выясняем, какой эффект даёт использование словарей *КВИ* наряду со словарями *ВИ*. Мы ожидаем увеличения коэффициента покрытия, но платой за это будет появление ошибок дешифровки.

Коэффициент покрытия песнопения цепочками знамен из словарей *ВИ* и *КВИ* подсчитывается следующим образом. Выделяем в песнопении все цепочки, соответствующие вхождению *ВИ* и *КВИ*. Подсчитываем суммарное число позиций, покрываемых этими цепочками в песнопении. Каждая позиция учитывается лишь *один раз*. Тем самым устраняется эффект вложенности или перекрываемости отдельных *ВИ* и *КВИ*. Кроме того, считаем, что *КВИ* лишь тогда вносит вклад в покрытие, когда осуществляемая с его помощью нотолинейная реконструкция с привлечением только доминирующей интерпретации даёт правильный результат. С учётом сказанного коэффициент покрытия песнопения цепочками из словарей *ВИ* и *КВИ* определяем как отношение $k = n / N$, где N – число знамен в песнопении (его длина). Схема покрытия фрагмента (первые 15 знамен) песнопения №10 из Октоиха 619/647 *ВИ* и *КВИ* длины $L = 1 \div 8$ иллюстрируется на рис. 1. Знамена, составляющие *ВИ*, обозначаются символом '=', а *КВИ* – символом '#'. Если знамя покрыто и *ВИ*, и *КВИ*, используется символ '&'. Цепочки, покрытые *КВИ*, но интерпретированные с ошибкой, обозначаются символом '?'.

L	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1		#		#		#		?					=		
2		?	=	&	&	=	?	?		#	#	&	=		
3			=	=	=	=	?	?	?		#	#	#	#	#
4			=	=	=	=					#	#	#	#	#
5											#	#	#	#	#
6										#	#	#	#	#	#
7								#	#	#	#	#	#	#	#
8								#	#	#	#	#	#	#	#
итог		#	=	=	=	=	?	#	#	#	#	#	=	=	#

Рис. 1. Схема покрытия фрагмента песнопения цепочками длины L из словарей *ВИ* и *КВИ*.

Нетрудно видеть, что в покрытии текста задействованы цепочки длины L от 1 до 8 знамен. При $L = 1$, например, используются *КВИ* ,  и , а также *ВИ* . Знамя "столица" () также выделено в качестве *КВИ* и интерпретировано в соответствии с доминирующим распевом как е2. Фактически же в двознаменнике стоит "d2", поэтому имеет место ошибка, помеченная знаком '?'. Эта ошибка исправляется лишь при $L = 8$ (в соответствующем *КВИ* "столица" стоит в самом начале). Знамя , с которого начинается рассматриваемый фрагмент песнопения, остался неидентифицированным. Он не покрыт ни одной из цепочек словарей *ВИ* и *КВИ*.

В последней строке рисунка приведено итоговое покрытие фрагмента, сформированное путём выбора наиболее сильного кандидата из столбца "решений", получаемых по поводу каждого знамени при разных значениях L . Порядок предпочтения таков: '=' > '&' > '#' > '?'.

Коэффициенты покрытия отдельных песнопений в каждом гласе меняются в широких пределах. Это обусловлено, по-видимому, жанровой спецификой. Ниже мы приводим средние по всем песнопениям гласа коэффициенты покрытия \bar{k} для той подборки, по которой происходило формирование словарей *ВИ* и *КВИ*. Рассматриваются три случая: а) покрытие

проводилось только с использованием *ВИ* ($F_{nop} = 3$); б) использовались *ВИ* и *КВИ* ($F_{nop} = 3; F_{dom}/F > 1/2$); в) повышен порог в пункте "б" при определении доминирующей интерпретации ($F_{nop} = 3; F_{dom}/F \geq 2/3$). С учётом оговоренных условий получены следующие результаты:

Гласы		1	2	3	4	5	6	7	8
ВИ ("а")	\bar{k}	0.640	0.487	0.619	0.677	0.626	0.510	0.799	0.619
ВИ+КВИ("б")	\bar{k}	0.826	0.777	0.786	0.833	0.825	0.772	0.890	0.807
ВИ+КВИ("в")	\bar{k}	0.805	0.722	0.742	0.804	0.773	0.727	0.874	0.781

Две последние строки подтверждают целесообразность использования *КВИ* для нотолинейной реконструкции знаменного распева: коэффициент покрытия по отдельным гласам возрастает на $15 \div 20\%$. Однако порядка $6 \div 10\%$ знамен реконструируются с ошибкой (в коэффициент покрытия эти ошибки не входят). Повышение порога с $1/2$ до $2/3$ при определении доминирующей интерпретации ожидаемо понижает коэффициент покрытия.

Более реалистичная оценка может быть получена в режиме "скользящего контроля". Из ("обучающей") подборки песнопений гласа удаляется по очереди одно песнопение («контрольное»), а по оставшимся строятся словари *ВИ* и *КВИ*. Подсчитывается коэффициент покрытия контрольного песнопения выделенными *ВИ* и *КВИ*. Затем песнопение возвращается в обучающую подборку, а в качестве «контрольного» рассматривается следующее песнопение, не анализировавшееся ранее. Процесс повторяется до исчерпания всех песнопений гласа. Средние значения коэффициента покрытия \bar{k} по каждому гласу двознаменника 619/647, полученные в режиме скользящего контроля, приведены ниже:

Глас		1	2	3	4	5	6	7	8
ВИ ("а")	\bar{k}	0.556	0.384	0.511	0.508	0.562	0.378	0.698	0.538
ВИ+КВИ("б")	\bar{k}	0.730	0.654	0.705	0.720	0.753	0.665	0.815	0.709
ВИ+КВИ("в")	\bar{k}	0.696	0.577	0.660	0.675	0.697	0.595	0.786	0.685

Можно видеть, что коэффициенты покрытия снижаются по сравнению с предыдущим экспериментом, что объяснимо, так как для дешифровки каждый раз предъявляется песнопение, не участвовавшее в построении словарей *ВИ* и *КВИ*. Тем не менее, добавление *КВИ* к *ВИ* значительно повышает коэффициент покрытия и в этом случае. Величина ошибки колеблется для случая "б" в диапазоне $9 \div 13\%$, а для случая "в" – в диапазоне $5 \div 9\%$.

Ещё один эксперимент по оцениванию покрываемости песнопений цепочками из словарей *ВИ* и *КВИ* был проведён с привлечением в качестве контрольного материала песнопений Октоиха 618/644, тогда как словари *ВИ* и *КВИ* строились на основе Октоиха 619/647. Ниже приведены средние значения коэффициентов покрытия в этом случае:

Глас		1	2	3	4	5	6	7	8
ВИ ("а")	\bar{k}	0.492	0.326	0.411	0.372	0.422	0.218	0.542	0.459
ВИ+КВИ("б")	\bar{k}	0.710	0.639	0.595	0.645	0.629	0.558	0.718	0.612
ВИ+КВИ("в")	\bar{k}	0.677	0.567	0.530	0.571	0.577	0.474	0.663	0.578

Эти результаты показывают, что имеет место дальнейшее снижение показателей покрываемости даже по сравнению с вариантом скользящего контроля. Более того, существенно возрастает число ошибок нотолинейной реконструкции, которое меняется в пределах от 19 до 26% с некоторым превалированием для случая "б". Это означает, что Октоихи заметно отличаются друг от друга на знаменном и нотолинейном уровне. Существенным может оказаться недостаточный объём обучающих подборок, различие в датировках Октоихов и (в пределах одного столетия) и подборе песнопений по каждому гласу.

6 Обсуждение результатов.

В работе использована самая простая, но не лучшая в плане допускаемых ошибок стратегия формирования и, соответственно, использования словаря *КВИ* — по наличию доминирующей интерпретации. Большую ошибку при этом дают высокочастотные знамена (типа "стопицы"), с небольшим запасом преодолевающие порог по доминированию ($1/2$ или

2/3). Существуют определённые резервы для снижения числа ошибок, требующие более "изолированного" подхода к формированию словаря *КВИ*. В частности, следует обратить внимание на характер вариативности тех цепочек, которые отбираются в словарь *КВИ* в соответствии с наличием доминирующей интерпретации. Те интерпретации, которые не являются доминирующими и характеризуются как ошибки, на самом деле далеко не случайны. Довольно часто они представляют собой *секвентные переносы доминирующей интерпретации*, которые относительно легко учесть и перевести из разряда ошибок в категорию "допустимого" варьирования.

Рассмотрим в качестве примера триграмму f4g4*a2*g4f4 , встретившуюся в первом глазе 17 раз с двумя интерпретациями f4g4*a2*g4f4 – 11 раз и e4f4*g2*f4e4 – 6 раз. При пороге доминирования 2/3 эта триграмма не преодолевает порог и выбывает из рассмотрения. При пороге 1/2 она преодолевает барьер и попадает в словарь *КВИ*, однако в 6 случаях из 17 она будет интерпретирована как f4g4*a2*g4f4 , что эквивалентно ошибке. Тем не менее, если учесть, что вторая интерпретация является секвентным переносом первой на ступень вниз и отразить эту информацию в словаре, ошибок можно избежать и даже внести дополнительный вклад в покрытие.

Более интересны случаи, когда ни одна из интерпретаций не является доминирующей, но все они являются секвентными переносами по отношению друг к другу. Учёт этого фактора позволяет объединить их в одну группу, где инвариантом является интервально-длительностная структура, и включить эту группу в словарь *КВИ*.

Довольно часто в цепочке знамен длины $L \geq 2$ варьируемым является лишь одно знамя. При этом все интерпретации цепочки, отличные от доминирующей, классифицируются как ошибки и помечаются знаком "?", хотя $(L - 1)$ знамя интерпретировано верно.

Количество подобных соображений, учёт которых может способствовать уменьшению числа ошибок дешифровки, легко увеличить. Все они основаны на детальном изучении вариативности знаменного распева. Более подробно вопросы разделения случайных ошибок и допустимого варьирования рассмотрены нами в [9]

Литература

- [1] Бражников М.В. Древнерусская теория музыки, Л.: изд. «Музыка», 1972, – 422 С.
- [2] 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.
- [3] 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.
- [4] Металлов В.М. Осмогласие знаменного распева (сборник нотолинейных попевок), с.1–50, М.: –1899 г.
- [5] Бражников М.В. Лица и фиты знаменного распева, Л.: изд. «Музыка», 1984, – 302 С.
- [6] Бахмутова И.В., Гусев В.Д., Титкова Т.Н. Создание электронной азбуки знаменного распева на основе анализа двознаменников // *Древнерусское песнопение. Пути во времени. По материалам международной научной конференции "Бражниковские чтения" 2008-2009 годов*, - т.4, - Изд-во Политехнического университета, г.С.-Петербург, - 2010 – с.99-108.
- [7] Бахмутова И.В., Гусев В.Д., Титкова Т.Н. Компьютерный поиск инвариантных структурных единиц знаменного распева, // *Проблемы музыкальной науки*, т. 1(8). – Уфа, Гидем, 2011, с.20-24.
- [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] Бахмутова И.В., Гусев В.Д., Титкова Т.Н. Автоматическое выявление разночтений и потенциально возможных ошибок в текстах двознаменников // *Проблемы музыкальной науки*, т. 1(12). – Уфа, Гидем, 2013, с.22-26.