

Л-граммные азбуки для дешифровки знаменных песнопений¹

Бахмутова И.В., Гусев В.Д., Титкова Т.Н.

Статья поступила 21 августа 1998 г.

г. Новосибирск

Рассматривается дешифровочная по своей природе и нерешенная в общем случае задача перевода древнерусских знаменных песнопений в современную нотолинейную форму. Важным элементом ее является выбор исходных единиц для дешифровки. Дошедшие до нас дешифровочные таблицы содержат весьма ограниченный ассортимент структурных единиц и ориентированы на дешифровку лишь пометных рукописей позднего периода (XVII–XVIII в.в.)

Учитывая, что язык знаменного распева является контекстно-зависимым, предлагается на основе двознаменников (песнопений, представленных параллельно в знаменной и нотолинейной форме) создать компьютерные дешифровочные таблицы нового типа, в которых каждое знамя представлено со всевозможными лево- и правосторонними контекстами регулируемой длины. Описываются критерии и алгоритмы отбора наиболее информативных (в плане дешифровки) единиц из полученных таблиц.

Введение





Древнерусские церковные песнопения XII–XVII веков, зафиксированные в многочисленных певческих рукописях, чаще всего представлены в знаменной форме записи, являвшейся ведущей в системе музыкальной письменности того времени. Знамена – специальные знаки, служащие для передачи музыкальных звуков. При переводе их в современную нотолинейную форму они интерпретируются цепочками нот переменной длины (обычно от 1 до 5 нотных знаков).

Правила знаменного распева передавались из поколения в поколение в устной форме. Лишь в XV–XVI веках появились краткие музыкальные руководства – азбуки, содержащие начертания, наименования и весьма приблизительные указания на способ исполнения отдельных знамен [1]. Азбуки не заменяли, а лишь дополняли сложившуюся традицию устного обучения. Многие, казавшиеся очевидными для того времени факты, в них либо не оговаривались, либо снабжались пояснениями типа: такое–то знамя исполнять “по обычаю”. Как отмечается в [2], “как раз этот обычай и не может быть восстановлен”.

Знаменный распев регламентируется системой *осмогласия*, в соответствии с которой певческие книги разбиваются на 8 самостоятельных частей (*гласов*), отличающихся друг от друга составом, мелодическим содержанием и частотой использования *попевок* – элементарных интонационных единиц древнерусской музыки. В древнегреческой системе осмогласия понятие гласа ассоциировалось с ладом (ионийский, дорийский и т.д. – всего 8 ладов). Несмотря на византийские истоки знаменного распева, русская система осмогласия существенно отличается от древнегреческой. В частности, понятие лада деформировалось, а средством мелодической характеристики гласа стала выступать система попевок.

¹ Работа выполнена в рамках проекта № 96-04-06258, поддержанного Российским Гуманитарным Научным Фондом, **опубликована в:** «Сибирский журнал индустриальной математики». – Т.1, № 2, Новосибирск. 1998. – С.51–66.

Под *дешифровкой* понимается процесс перевода песнопений из знаменной формы записи в современную нотопевицкую. Первые дешифровки были сделаны в XVII веке, однако этот процесс не получил должного (и своевременного) теоретического осмысления и развития. Как следствие, певческие книги XVI века и более раннего периода остаются практически нечитаемыми. Тем не менее именно в XVII веке появился ряд *двознаменников* – музыкальных текстов, записанных параллельно в знаменной и нотопевицкой форме. Они играют роль *билингв*, наличие которых существенно облегчает процесс дешифровки.

Основная трудность дешифровки, на наш взгляд, связана с тем, что язык знаменного распева является *контекстно-зависимым*. Это проявляется в том, что интерпретация отдельных знамен зависит от типа элементарных структурных единиц, в которые они входят, места расположения этих единиц в мелодии, гласовой принадлежности и ряда других факторов, что обуславливает *многозначность* соответствий “знамя–нота” и “нота – знамя”. Например, знамя  (“статья с запятой”), дважды входящее в состав попевки “кавычка” из гласа б:   , в третьей позиции интерпретируется цепочкой из трех нот *e2d2c4*, а в пятой позиции – одной целой нотой *d1* (система обозначений и первоисточник указаны в разделе 1).

Трудности интерпретации усугубляются *отсутствием* объективного *критерия правильности* дешифровки, *неполнотой* (а иногда и *противоречивостью*) дошедших до нас дешифровочных таблиц (азбук “знамя–нота” и сборников попевок), слабым представлением о *системе осмогласия* и *эволюции* знаменного распева.

Целью работы является *создание дешифровочных таблиц нового типа*, названных авторами *L-граммными азбуками*. В отличие от традиционно используемых таблиц, обобщающих разрозненные данные из рукописных *музыкально-теоретических* руководств, предлагаемые таблицы создаются на основе анализа реальных музыкальных текстов, представленных в форме *двознаменников*. Это позволяет снабдить их количественной информацией, полезной для дешифровки. *L-граммная* форма представления азбук обеспечивает свободу выбора *единиц дешифровки* и существенно расширяет их спектр – от отдельных знамен до комбинаций различных структурных единиц.

1. Терминология, система обозначений, исходные данные.

Используемые на практике дешифровочные таблицы имеют форму *двознаменников* и отличаются по типам интерпретируемых единиц. В *азбуках* приводятся нотопевицкие интерпретации отдельных знамен. Алфавит знамен в дальнейшем будем обозначать через Σ , а число элементов алфавита – через $|\Sigma|$.

В *кокизниках* представлены более крупные единицы – попевки, составляющие мелодическую основу знаменного распева. В *фитниках* собраны специфические (“тайнозамкненные”) мелодические обороты – лица и фиты, служащие в основном для украшения знаменного распева. Под “*тайнозамкненностью*” понимается “графический прием зашифровки напева посредством такого условного сочетания знамен, которое не образует этого напева в случае исполнения знамен в этом же последовании, но согласно их обычному певческому значению” [2].




Изъятие лиц и фит из знаменного распева обедняет напев, но в отличие от попевок не разрушает его. Мы уделим основное внимание созданию таблиц для интерпретации знамен, попевок и некоторых других типов структурных единиц, оставляя в стороне лица и фиты.

Анализ и систематизацию значительного числа кокизников провела А.Н.Кручинина [3]. Все попевки по типам *кадансов* (завершений) разделены на 24 семейства, каждое из которых имеет свою первооснову (или “«архетип»”). Попевка образуется путем добавления слева к 3–элементному архетипу (или его производному) цепочки знамен, называемой “*подводом*” (примерно из 3–4 знаков).

Введению нотной интерпретации предшествовала система *помет* – специальных киноварных (по большей части) знаков, которые проставлялись у отдельных знамен и указывали на способ их исполнения. Различают два типа помет: *степенные* – осуществляющие звуковысотную привязку знамен, и *указательные*, определяющие характер исполнения знамен в определенном контексте. Последние проставляются реже, но изменяют мелодическую и ритмическую характеристики знамен. Известные (и очень немногочисленные) примеры дешифровок относятся к пометным рукописям XVII века (см. для иллюстрации [4]), т.е. представляют собой частные решения общей задачи, и в деталях не описаны.

В приводимых далее нотных иллюстрациях используется следующая система обозначений. Звуки первой октавы помечены заглавными буквами: *G* – “соль”, *A* – “ля”, *H* – “си”, звуки второй – прописными: *c* – “до”, *d* – “ре”, *e* – “ми”, *f* – “фа”, *g* – “соль”, *a* – “ля”, *b* – “си–бемоль”. Стоящие рядом с буквами справа цифры обозначают длительности нот: 1 – целая, 2 – половинная, 4 – четвертная, 8 – восьмая, 2. – половинная с точкой и т.д.} Знак * используется в качестве разделителя между нотолинейными интерпретациями разных знамен.

Для выявления степени отличия различных нотолинейных интерпретаций одного знамени (или цепочки знамен) удобно использовать интервальное представление. Под *интервалом* понимается число ступеней звукоряда между двумя соседними звуками. Формально, если $(h_1 d_1)(h_2 d_2) \dots (h_l d_l)$ – цепочка нотных знаков, где h_k – высота k – го звука, d_k – его длительность, $1 \leq k \leq l$, l – длина цепочки ($l \geq 2$), то интервальное представление цепочки имеет вид $I_1 I_2 \dots I_{l-1}$, где $I_k = h_{k+1} \ominus h_k$, $1 \leq k \leq l-1$; \ominus – символ специальной операции вычитания. Поскольку величина интервала может быть положительной (движение вверх), отрицательной (движение вниз) или равной 0 (повтор звука на той же высоте), используем для обозначения I запись $(|I_k|+)$ в первом случае, $(|I_k|-)$ – во втором и $(0+)$ – в третьем, хотя при $(|I|=0)$ знак “+” ставится условно (в качестве разделителя соседних цифр).

Нетрудно видеть, что нотные цепочки, отличающиеся лишь секвентным переносом, имеют одинаковую интервальную структуру. Так, цепочки *H4c4d4c4* и *d4e4f4e2* имеют структуру (1+1+1–), но отличаются ритмически ( и , тогда как цепочки *H2H2G1* и *e2e2c1* эквивалентны и в интервальном представлении (0+2–), и в ритмическом ().

Исходным материалом для нас послужила одна из наиболее распространенных богослужебных певческих книг – “Октоих”. Анализируются два “Октоиха”. Первый принадлежит собранию Кирилло–Белозерского монастыря (КБ), содержит свыше 400

песнопений, пометный, (ОР РНБ им. Салтыкова–Щедрина, г.Санкт–Петербург, Кирилло–Белозерское собрание., N632/889, XVII век.). Второй – из собрания Соловецкого монастыря (СО), содержит более 200 песнопений, представлен в форме *двознаменника*, знамена снабжены пометами (ОР РНБ, Соловецкое собрание, N619/647, XVII в.).

2. Краткая характеристика существующих дешифровочных таблиц

Прототипом для используемых в настоящее время дешифровочных таблиц послужили многочисленные певческие азбуки Древней Руси. Их эволюция отражена в [1]. Эти азбуки носят описательный характер и подразделяются на азбуки–перечисления, представляющие алфавит знамен, и азбуки–толкования, объясняющие (весьма приблизительно) характер исполнения каждого знамени (“стопица с очком – поотдати голосом”, “громную стрелу с сорочьей ногою, подержав, дважды голкни и светло закинь” т.п.). Данные азбуки затруднительно использовать непосредственно для целей дешифровки, но они являются единственными ориентирами при анализе ранних (беспометных) рукописей.

Наибольшей известностью в дешифровочной практике пользуются более поздние азбуки, созданные в конце XIX века В.М.Металловым, Д.В.Разумовским и С.В.Смоленским. Они отражают личный опыт авторов и являются результатом компиляции данных, почерпнутых из более ранних (и не всегда указываемых) музыкально–теоретических руководств и рядовых певческих рукописей. Как следствие, эти азбуки обладают разной степенью полноты и порою противоречивы [5].

Отметим основные отличительные особенности указанных азбук:

1) в качестве единиц дешифровки фигурируют отдельные знамена, попевки, лица и фиты;

2) каждой единице соответствует ее нотолинейный эквивалент с точностью до звуковысотной привязки;

3) знамена снабжены указательными пометами, призванными подчеркнуть контекстные различия в интерпретации знамен. Одно и то же знамя с разными указательными пометами рассматривается как совокупность самостоятельных элементов алфавита. Неявно предполагается, что эти элементы отличаются по своей функциональной нагрузке, т.е. входят в состав различных структурных единиц.

Поскольку обсуждаются пометные азбуки, созданные в XIX веке, их применимость для дешифровки пометных же рукописей XVIII и особенно XVII века уже требует специального обоснования. Что же касается дешифровки беспометных рукописей (а именно эта задача является наиболее сложной), то пометные азбуки малопригодны для этого по следующим причинам:

1) связь между используемой при знамени указательной пометой и контекстом не просматривается. А только этот факт и мог бы быть использован при дешифровке беспометных текстов. Размер контекста должен быть достаточным для однозначной интерпретации знамени;

2) ассортимент дешифровочных единиц, представленных в азбуках, весьма ограничен. Анализ реальных песнопений указывает на наличие промежуточных градаций между знаменем и попевкой (подобных морфемам в словах естественного языка). Наблюдаются также структурные единицы, отличные от попевок (тандемные повторы, устойчивые начальные обороты, гласоспецифические цепочки и т.п.);

3) малоинформативными представляются звуковысотные пометы (степенные и признаки), предоставляемые у знамен в азбуках. В реальных песнопениях каждому знамени соответствует целый спектр возможных звуковысотных привязок и проблема как раз и заключается в выборе нужной привязки в зависимости от контекста.

Указанные соображения (их список может быть продолжен) свидетельствуют о том, что ориентация на дешифровку *беспомянутых* рукописей требует создания дешифровочных таблиц *нового* типа. Каждое знамя должно быть представлено с множеством допустимых лево- и правосторонних *контекстов* регулируемого размера. Для каждого знамени (или цепочки знамен) должен быть указан *полный спектр* его возможных *нотоподобных интерпретаций*, отличающихся в общем случае интервально-ритмической структурой, а в простейшем случае – лишь звуковысотной привязкой. Все интерпретации снабжаются информацией о *частоте* их использования в данном гласе, что позволяет составить представление о “нормативной” и “ненормативной” трактовке каждого знамени.

Ниже излагаются основы предлагаемого решения.

3. Формирование азбук по двознаменникам

В основу формирования новых азбук положены тексты рядовых певческих рукописей, представленные в форме двознаменников. Азбуки составляются для каждого гласа в отдельности. Элементами азбук выступают цепочки знамен переменной длины. Следуя Шеннону [6], будем называть их L -граммами (L – длина цепочки). L -грамме x_L ставится в соответствие множество ее всевозможных нотоподобных интерпретаций в данном гласе – $\text{Int}(x_L)$, где для каждой интерпретации $\text{int}(x_L)$ указывается суммарная частота ее встречаемости $F(\text{int}(x_L))$ во всех песнопениях гласа. Очевидно, что

$$F(x_L) = \sum_{\text{Int}(x_L)} F(\text{int}(x_L)).$$

Однако, если рассматривать нотоподобную цепочку $\text{int}(x_L)$ в отрыве от цепочки знамен x_L , то частота ее появления в гласе может оказаться выше, чем $F(\text{int}(x_L))$, ввиду наличия других цепочек знамен с той же нотоподобной интерпретацией. Иными словами, многозначным является не только соответствие “знамя–нота”, но и “нота–знамя”. Если $|\text{Int}(x_L)|=1$, то цепочка x_L является *однозначно* интерпретируемой в данном гласе.

В азбуки включаются *все повторяющиеся* в данном гласе цепочки знамен длины $1, 2, \dots, L_{\max}$, где L_{\max} – длина максимального повтора в гласе. Они дополняются *однократно* встречающимися цепочками, которые становятся кратными при редуцировании последнего символа. Их можно считать “однократными в первом поколении”, поскольку все последующие их расширения также будут иметь частоту 1. Такого рода цепочки являются однозначными (и кратчайшими) идентификаторами своих позиций [7] в том смысле, что любая их префиксная подцепочка уже обязательно встретится в какой-либо еще позиции текста. Под текстом в данном случае понимается конкатенация $T = T_1 \vee T_2 \vee \dots \vee T_m$ всех песнопений гласа, где \vee – разделитель между песнопениями. В азбуки не включаются цепочки, содержащие разделитель.

Множество цепочек $\Phi(T)$, выделяемое из текста T , представляется в виде совокупности подмножеств $\{\Phi_L(T)\}$, каждое из которых содержит лишь цепочки

фиксированной длины L ($L = 1, 2, \dots, L_{\max} + 1$). Подмножество $\Phi_L(T)$ будем называть *L-граммной характеристикой* (или *L-граммной азбукой*) текста T . Цепочки, составляющие $\Phi_L(T)$, представлены в виде: а) частотного и б) лексикографического упорядочения.

Вычисление L -граммных характеристик ведется итеративно по L [8]. Каждая итерация связана с однократным просмотром текста. Для каждой L -граммы текста с помощью процедуры хеширования определяется адрес ее счетчика в оперативной памяти, где хранится информация обо всех вхождениях данной L -граммы в текст T . Отображение " L -грамма \rightarrow адрес" осуществляется с помощью модульной хеш-функции:

$$h(K(x_L)) = K(x_L) \bmod M, \quad (1)$$

где $K(x_L)$ – числовой код L -граммы x_L , M – размер расстановочного поля, определяемый мощностью словаря L -грамм, который заведомо не превышает длины текста T (точнее, $\sum_{i=1}^m (N_i - L + 1)$, где N_i – длина песнопения T_i). Отображение (1) гарантирует попадание одинаковых L -грамм в один и тот же счетчик, но допускает и возможность наложения разных L -грамм, т.е. возникновение ситуаций, когда $h(x_L) = h(y_L)$ при $x_L \neq y_L$. Такого рода наложения устраняются с помощью стандартной списковой схемы [8].

Трудоёмкость каждой итерации может быть существенно снижена, если:

а) использовать рекуррентную схему вычисления значений хеш-функции (1), основанную на "зацепленности" соседних L -грамм текста (i -я и $i+1$ -я L -граммы текста – $T[i:i+L-1]$ и $T[i+1:i+L]$ – имеют общее $(L-1)$ – элементное ядро $T[i+1:i+L-1]$). Благодаря этому трудоёмкость хеширования одной L -граммы не растёт с увеличением L [9];

б) учесть тот фактор, что если какая-нибудь $(L-k)$ -грамма x_{L-k} ($k < L$) представлена в $\Phi_{L-k}(T)$ и $F(x_{L-k}) = 1$, то всевозможные ее право- и левосторонние расширения уже отсутствуют в характеристиках более высокого порядка, т.к. их частота заведомо будет равна 1. Применительно к характеристикам L -го порядка это означает, что существует кластер из k подряд расположенных L -грамм (включающих в себя $(L-k)$ -грамму с единичной частотой), которые не надо хешировать. Простой механизм отслеживания таких кластеров позволяет с увеличением L хешировать все меньшее число L -грамм.

Из а) и б) вытекает, что трудоёмкость одной итерации, связанной с получением характеристики L -го порядка, составляет в наихудшем случае $O(N)$, где $N = \sum N_i$ – суммарная длина всех песнопений гласа. Из б) следует также, что нет смысла продолжать итерации вплоть до значения $L_{\max} + 1$. Достаточно ограничиться некоторым значением $L^* < L_{\max}$ таким, что число подлежащих хешированию L^* -грамм станет небольшим по сравнению с длиной текста и (или) их частоты не снизятся до значений $F=2,3$. Соответствующие L^* -граммы тогда могут быть расширены непосредственно по тексту для выявления границ повторов. С учетом вышесказанного суммарная трудоёмкость построения азбук для одного гласа составляет

$O(L^* \cdot N)$, где $L^* = c \cdot L_{\max}$, $0 < c < 1$ (в наших экспериментах константа c была близка к $2/3$).

В заключение данного раздела сделаем несколько замечаний.

1) Идеология L -граммного представления текста восходит к работам Шеннона по теории информации и криптографии (см., в частности, [6]). Она наиболее эффективна применительно к *слитным неструктурированным* текстам, характерным для криптографических приложений. Таковыми являются и знаменитые песнопения.

2) Мы не ограничиваем себя относительно небольшими значениями L , как это имеет место в большинстве приложений, а используем *полный спектр повторов* $\Phi(T) = \{\Phi_1(T), \Phi_2(T), \dots, \Phi_{L_{\max}+1}(T)\}$, поскольку только с его помощью могут быть выявлены структурные единицы на разных иерархических уровнях (см. п. 4.5).

Предлагаемая нами структура данных – полный частотный спектр повторов $\Phi(T)$ – обладает теми же функциональными возможностями, что и префиксное дерево (конструкция, предложенная Вайнером [7]), внутренние узлы которого дополнены информацией о частоте встречаемости в тексте заканчивающихся в них подцепочек. В то же время наш алгоритм вычисления $\Phi(T)$ (рекуррентное хеширование) много проще в логическом отношении алгоритма Вайнера и, как следствие, сопоставим с ним по трудоемкости на текстах достаточно большой длины, несмотря на худшие показатели в асимптотике.

3) Характеристика $\Phi_{L+1}(T)$ содержит информацию о всех правосторонних расширениях цепочек из $\Phi_L(T)$, имеющих неединичную частоту. Так, если цепочка $x_L \in \Phi_L(T)$, не является конечной ни в одном из T_i ($i = 1 \dots m$), составляющих текст T , и $F(x_L) \geq 2$, то в $\Phi_{L+1}(T)$ обязательно представлены цепочки вида $x_L a$, где a – произвольный элемент алфавита Σ , причем

$$F(x_L) = \sum_{a \in \Sigma} F(x_L a) \quad (2)$$

При лексикографическом упорядочении элементов из $\Phi_{L+1}(T)$ цепочки вида $x_L a$ будут локализованы в одном месте.

Для удобства выделения структурных единиц в соответствии с процедурой, описанной ниже, нам понадобится информация о всевозможных левосторонних расширениях каждой цепочки x_L . Соответствующие $(L+1)$ -граммы, имеющие вид ax_L , где $a \in \Sigma$, представлены в $\Phi_{L+1}(T)$ и для них выполняется соотношение, аналогичное (2):

$$F(x_L) = \sum_{a \in \Sigma} F(ax_L),$$

при условии, что x_L не является начальной L -граммой ни в одном из текстов T_i ($i = 1, \dots, m$). Однако при лексикографическом упорядочении элементов из $\Phi_{L+1}(T)$ цепочки вида ax_L уже не будут локализованы в одном месте, что затрудняет их поиск. Выходом из положения является инвертирование цепочек из $\Phi_{L+1}(T)$, т.е. запись их в обратном порядке (справа налево), с последующей лексикографической сортировкой. Такого рода характеристику будем называть *инвертированной* и обозначать $\Phi_{L+1}''(T)$. В $\Phi_{L+1}''(T)$ цепочки вида ax_L из $\Phi_{L+1}(T)$ будут представлены в обратном порядке, но локализованы в одном месте. Формально $\Phi_{L+1}''(T) = \Phi_{L+1}(T'')$, где T'' – инвертированный текст T . Инвертированные характеристики будем использовать при отслеживании всевозможных левосторонних расширений заданной цепочки.

4. Выделение структурных единиц знаменного текста с помощью *L*–граммных азбук.




Цепочки знамен, представленные в *L*–граммных азбуках, обладают разной степенью информативности. Для целей дешифровки следует отобрать наиболее информативные из них. Ниже описаны принципы, которые могут быть положены в основу подобного отбора.

4.1. Алфавит знамен. Он должен быть представлен полностью, поэтому характеристика *первого порядка* информативна “по определению”. Форма представления знамен в ней иллюстрируется в таблице на примере гласов 1 и 6 двознаменного “Октоиха” (полные характеристики представлены в [10], [11])². Приводимые ниже комментарии целиком основаны на той

дополнительной информации, которая содержится в *L*–граммных азбуках по сравнению с традиционными. В первую очередь речь идет о *количественном* учете всех форм многозначности, включая компоненту, связанную с изменением звуковысотной привязки.

Знамена в строках 1 и 2 (“крыж” и “статья простая”) – кадансовые по своей природе – иллюстрируют систему *опорных звуков* в каждом гласе. Поскольку в конце песнопений обычно стоит “крыж” (†), его нотолинейные эквиваленты определяют главные опорные звуки (“устои”) каждого гласа. В гласе 1 их два, примерно равночастотные (*G* – “соль” 1-й октавы и *d* – “ре” 2-й октавы). В гласе 6 в качестве “устоя” фигурирует лишь один звуковысотный уровень – “ре” 2-й октавы. Наличие нескольких “устоев” в одном гласе свидетельствует о том, что в системе русского осмогласия понятие гласа уже не ассоциируется с понятием лада, как это характерно для греческого осмогласия.

Знамена в строках 3 и 4 (“стопица” и “стрела простая”) согласно беспометным азбукам–перечислениям определяют звуковысотный уровень “строки” – своеобразного *начала отсчета*, по отношению к которому (выше или ниже) задается высотный уровень других знамен. Наличие многих нотолинейных эквивалентов у этих знамен и несовпадение высот у главных (наиболее частых) интерпретаций “стопицы” и “стрелы” в каждом гласе свидетельствует о размывании понятия “строки” в ходе эволюции знаменного распева.

Знамя  (“стрела светлая” – см. строку 5 в таблице) *однозначно* интерпретируется в гласе 1 и “почти однозначно” в гласе 6. Более того, ее интерпретации в обоих гласах совпадают. Однозначно интерпретируемые знамена особенно важны для дешифровки, поскольку образуют систему *звуковысотных инвариантов* в каждом гласе. В таблице представлены и другие примеры такого рода инвариантов: “стрела громная”  в первом гласе и “скамейца с точкой”  – в шестом.

Строки 7 и 8 таблицы демонстрируют *гласоспецифичность* отдельных знамен, т.е. употребимость их лишь в отдельных гласах. Такого рода знамена важны в плане определения гласовой принадлежности и понимания системы осмогласия в целом.

² В работах [10,11] по вине авторов вместо шифра 619/647 указан шифр 618/644.

Примеры интерпретации знамен в гласах 1 и 6 двознаменника (СО)

(F – частота знамени: частоты его нотолинейных эквивалентов указаны в скобках :


I – интервальная структура интерпретирующих цепочек длины 2 и выше)

N	Знамя	Глас 1			Глас 2		
		F	нотолнейный эквивалент	I-структура	F	нотолнейный эквивалент	I-структура
1	2	3	4	5	6	7	8
1		29	d1(15), G1(13), e1(1)		27	d1(27)	
2		138	d1(57), e1(48), A1(16), G1(11), d2(2), e2(1) ...		98	d1(71), c1(18), A1(6), H1(1), D2(1)	
3		230	e2(161), d2(33), c2(9), g2(7), f2(7), e4(5) ...		154	d2(54), f2(33), e2(29), c2(22), d4(9), f4(2) ...	
4		14	f1(12), e1(2)		31	c4d4e4(12), f1(10), e1(3) ...	1+1+
5		22	e4f4g2(22)	1+1+	44	e4f4g2(41), g4a4(1) ...	1+1+ 1+
6		22	A1(22)		4	c4H4c4d4(3), e4d4e2(1)	1-1+1+ 1-1+
7		21	e4f4d4(16), d4f4d4(2)	1+2- 2+2-	0		
8		0			43	d2c4(40), d4e4(1)	1- 1-
9		86	e1(41) c1(4), e2.d4(14), d2.c4(9) c4d4c4H4(14)	1- 1+1-1-	109	c4d4e4(41), d1(17), d2c2(13), f2.e4(10), d2.c4(8), c2H4c4(5) ...	1+1+ -, 1- 1- 1-1+




Строки 4, 6 и 9 иллюстрируют возможность принципиально разных интерпретаций знамени как в одном гласе, так и в разных. Здесь же можно наблюдать пример многозначности обратного соответствия – “нота–знамя”: цепочка *c4d4e4* из гласа 6 в одном случае представляет “стрелу простую” , а в другом – “статью закрытую малую”

Как правило, интерпретации одного знамени, отличающиеся по своей I-структуре, несут разную функциональную нагрузку. Так, интерпретация *c4d4e4* “статьи закрытой малой” в гласе 6 (см. Строку 9) встречается преимущественно в составе попевки “мережа” (срединный вариант). Если “мережа” завершает песнопение, то это же знамя

³ Многоточие в столбцах 4 и 7 означает, что указаны не все нотолинейные интерпретации. О суммарной частоте пропущенных комбинаций можно судить по разности между частотой знамени F (столбец 3 или 6) и суммой частот его нотолинейных эквивалентов (столбец 4 или 7).

интерпретируется как $d2c2$. В составе попевки “кулизма” знамя  интерпретируется целой нотой $d1$, а в одной из модификаций “грунки” – цепочкой $c2H4c4$. Как уже отмечалось выше, в традиционных азбуках эта многозначность *частично* снимается написанием при знамени различных указательных помет (правда, без связи их с функциональной нагрузкой). В данном же конкретном случае все вхождения “статьи закрытой малой” в знаменный текст не сопровождаются указательными пометами и для правильной интерпретации такого “беспометного” знамени в общем случае потребовалось бы установление по контексту типа структурной единицы, в состав которой оно входит. Для этих целей используется полный спектр L -граммных характеристик (см. п. 4.4).


4.2. L --граммные звуковысотные инварианты. В предыдущем подразделе мы продемонстрировали, что некоторые знамена *однозначно* интерпретируются в том или ином гласе и, как следствие, могут служить звуковысотными ориентирами при дешифровке беспометных рукописей. Свойство однозначной интерпретации свидетельствует о *высокой информативности* соответствующих знамен. Подтверждением может служить анализ других эволюционирующих систем (в частности, естественного и генетического языка), в которых фрагменты языковых единиц, не меняющиеся в ходе эволюции, считаются наиболее значимыми в функциональном отношении.

Однозначно интерпретируемых элементов (знамен) в гласовых характеристиках первого порядка встречается относительно немного. Однако в характеристиках порядка 2 и выше число однозначно интерпретируемых в данном гласе цепочек длины L ($L \geq 2$) растет. Это объясняется тем, что неоднозначно интерпретируемая цепочка x_L при расширении может стать однозначно интерпретируемой. Например, достаточно высокочастотные знамена  и  из гласа 1 имеют соответственно 5 и 11 интерпретаций, отличающихся в основном звуковысотной привязкой, однако их комбинация  (F=27) уже интерпретируется однозначно: $e2*f4e4$.

Наряду с эффектом увеличения с ростом L числа однозначно интерпретируемых цепочек существует и обратная тенденция. А именно, если цепочка x_L является однозначно интерпретируемой в гласе, то ее всевозможные (или отдельные) расширения типа $x_L a$ или $a x_L$ $a \in \Sigma$ уже могут не обладать этим свойством. Вычленив из структурной единицы максимальную цепочку, которая еще интерпретируется однозначно, мы выявляем тем самым наиболее значимую часть структурной единицы.

Исходя из вышесказанного, пополняем словарь информативных L -грамм, куда мы уже отнесли все элементы алфавита, цепочками однозначно интерпретируемых знамен длины $L \geq 2$, частота которых в гласе превышает некоторый порог. Последнее ограничение отсеивает L -граммы, однозначная интерпретируемость которых обусловлена их малочисленностью.

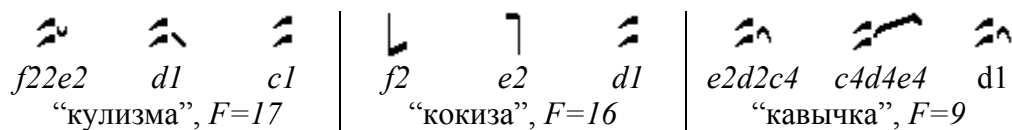
Приведем примеры однозначно интерпретируемых цепочек длины $L \geq 2$ из разных гласов.

Глас 1. Уже упоминавшаяся выше цепочка  расширяется, продолжая оставаться однозначно интерпретируемой, до значения $L=4$:

$$L \quad L \quad \text{[diagram]} \quad A \quad \leftrightarrow \quad e2 * e2 * f4 e4 * d2 \quad (F=16)$$

скачок $|I|=3+$). По совокупности указанных факторов (однозначная интерпретируемость, неявно выраженный каданс, наличие “параклита” и ему подобных знамен при расширении вправо) относим эту цепочку к числу самостоятельных структурных единиц, не фигурирующих однако в традиционных азбуках.

При $L=3$ выделяются однозначно интерпретируемые архетипы:



и ряд других.

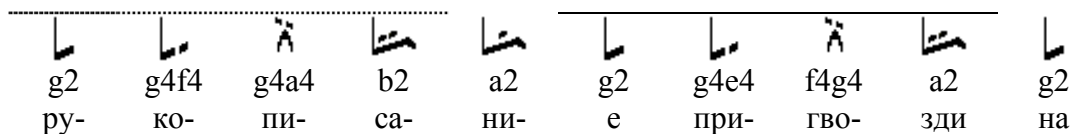
Приведенные примеры показывают, что однозначно интерпретируемые цепочки являются или самостоятельными структурными единицами (попевка), или подъединицами более низкого иерархического уровня (подвод, архетип, начальный фрагмент подраздела песнопения). Обнаружение их в беспометном знаменном тексте создает систему достаточно правдоподобных звуковысотных привязок, существенно облегчающих дешифровку.

4.3. Тандемные повторы. Другой способ выделения структурно значимых цепочек основан на привлечении информации о местах вхождения их в текст. Наиболее значимыми обычно являются цепочки, завершающие песнопение или самостоятельные его подразделы. Последние могут быть выделены, исходя из расстановки “параклитов” и “крыжей” в тексте песнопения, знаков препинания в стихотворном тексте и на основе ряда других признаков.

Специфическим типом позиционно выделяемых цепочек являются *тандемные повторы*. В простейшем случае речь идет о сериях, образуемых повторением одного и того же знамени. Наиболее часто встречаются серии “столиц” , соответствующие речитативным участкам, и серии из разновидностей крюков , , . Повторы остальных знамен встречаются редко и длины серий обычно не превышают 2.

В общем случае речь идет о тандемных повторениях цепочек длины $L \geq 2$, составленных из различных знамен. Такие танделы нередко встречаются в разводах (знаменных интерпретациях) лиц и фит, однако их немало и в рядовых знаменных текстах, лишенных элементов тайнозамкнутости. Элементом тандемного повтора, в частности, может служить и отдельная попевка.

Тандемные повторы следует считать самостоятельными структурными единицами, исходя хотя бы из того факта, что далеко не всегда повтор на знаменном уровне соответствует повтор на нотолинейном уровне, как в нижеследующем примере («Октоих», СО, глас 2.):



Это означает, что тандемные повторы подобно лицам и фитам могут нести в себе элемент тайнозамкнутости и, как следствие, требуют специального изучения. В традиционных азбуках структурные единицы подобного типа не рассматриваются.

4.4 Выделение структурных единиц на основе знаменного текста (частотно--позиционные критерии). Для выделения и интерпретации структурных единиц, представленных в разделах 4.1–4.3, использовались двознаменники. При дешифровке *беспомянутых* рукописей правильная интерпретация каждого знамени связана с определением его функциональной нагрузки, что в конечном итоге сводится к выявлению непосредственно по *знаменному тексту* структурной единицы, в состав которой входит знамя. Для этих целей могут быть использованы некоторые частотно–позиционные критерии достаточно общего вида.

Частотный критерий, начальная версия которого опубликована в [12], основан на предположении, что элементарные семантические единицы языка проявляют себя в виде устойчиво повторяющихся цепочек символов. Понятие устойчивости нуждается в уточнении.

Пусть $p = a_1 a_2 \dots a_n$ – произвольная цепочка текста T , составленная из элементов алфавита Σ . Для всевозможных префиксных подслов этого слова имеют место следующие соотношения:

$$F(a_1) \geq F(a_1 a_2) \geq \dots \geq F(a_1 a_2 \dots a_{n-1}) \geq F(p)$$

где $F(s)$ – частота встречаемости цепочки s в тексте T . Назовем цепочку p устойчивой при правостороннем расширении (или устойчивой справа), если:

1) существует $1 \leq i < n$ такое, что:

$$F(a_1 a_2 \dots a_i) \succ F(a_1 a_2 \dots a_i a_{i+1}) \succ \dots \succ F(p)$$

где знак \succ означает “не намного больше или равно” (это условие появления доминантной по частоте линии);

2) $F(p) > I$;

3) $F(p a_{n+1}) = F(a_1 a_2 \dots a_n a_{n+1}) \ll F(p)$, где $a_{n+1} \in \Sigma$, а $p a_{n+1}$ – произвольная цепочка текста T .

Первое условие означает, что как только формируемая цепочка $a_1 a_2 \dots a_i$ приобретает содержательный смысл, ее продолжение a_{i+1} легко предсказуемо, т.е. частота цепочки $a_1 a_2 \dots a_i a_{i+1}$ в тексте T доминирует над частотами цепочек вида $a_1 a_2 \dots a_i \alpha$ ($\alpha \in \Sigma, \alpha \neq a_{i+1}$), реализующих другие возможные продолжения. Слово “чемпион”, например, предсказуемо уже по цепочке “чемп”.


Условие 2 исключает цепочки с $F=I$, поскольку любые их расширения (на один и более символов) имеют ту же частоту и формально являются неделимыми. Условие 3 соответствует прерыванию доминантной по частоте линии, фиксируемой условиями 1 и 2 (слово “чемпион” продолжаемо уже разными способами: чемпионат, чемпионский, чемпионка и т.п.; поскольку каждое из продолжений содержит исходное слово, частота любого из них существенно меньше частоты исходного слова).

Аналогично, цепочку $p = a_1 a_2 \dots a_n$ текста T назовем устойчивой при левостороннем расширении (устойчивой “слева”), если существует $1 < j \leq n$, такое, что:

$$F(a_j a_{j+1} \dots a_n) \succ F(a_{j-1} a_j \dots a_n) \succ \dots \succ F(p), \quad F(p) > 1,$$

$$F(a_0 p) = F(a_0 a_1 a_2 \dots a_n) \ll F(p), \quad a_0 \in \Sigma,$$

а $a_0 p$ – произвольная цепочка текста T . Потенциальной *структурной единицей* будем считать цепочку текста, устойчивую как слева, так и справа. Это фрагмент, заключенный между левой и правой точками прерывания доминантной по частоте линии.

Алгоритм отыскания структурной единицы напоминает “качели”. Отправляясь от какого-либо элемента алфавита (или цепочки элементов), анализируем возможные продолжения в поисках участков стабилизации частот. Зафиксировав, например, правую границу, идем в обратном направлении в поисках левой границы. Возможны случаи, когда правая граница при первом проходе не фиксируется (не успевает проявить себя участок стабилизации частот по той, например, причине, что начальная цепочка слишком близко находилась от правой границы). Возвращаясь к исходной цепочке, осуществляем левостороннее расширение. Зафиксировав левую границу, вновь движемся от нее направо (второй проход) в поисках правой границы (отсюда термин “качели”). Проиллюстрируем вышесказанное на примере цепочки  ($l=2, F=17$) из гласа 8 (“Октоих”, КБ). Рассмотрим для простоты доминирующие по частоте правосторонние расширения этой цепочки и проследим за изменением частот $F(a_1)$, $F(a_1 a_2)$ и т.д.

$$\rightarrow \quad \begin{array}{cccc} a_1 & a_2 & a_3 & a_4 \\ \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \end{array} & \begin{array}{c} \text{---} \\ \text{---} \end{array} & \begin{array}{c} \text{---} \\ \text{---} \end{array} & \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \end{array} \\ 38 & 17 & 7 & 2 \end{array}$$


Здесь число, стоящее в нижней строке под символом a_i означает $F(a_1 a_2 \dots a_i)$. Нетрудно видеть, что при расширении цепочки вправо стабилизации частот не происходит. Вновь вернемся к исходной цепочке и попытаемся ее расширить справа налево, отслеживая частоты $F(a_2)$, $F(a_1 a_2)$, $F(a_0 a_1 a_2)$ и т.д.:

$$\begin{array}{c} a_{-2} \\ \begin{array}{c} \text{---} \\ \text{---} \end{array} \\ 5 \end{array} \quad \left| \quad \begin{array}{cccc} a_{-1} & a_0 & a_1 & a_2 \\ \begin{array}{c} \text{---} \\ \text{---} \end{array} & \begin{array}{c} \text{---} \\ \text{---} \end{array} & \begin{array}{c} \text{---} \\ \text{---} \end{array} & \begin{array}{c} \text{---} \\ \text{---} \end{array} \\ 12 & 12 & 17 & 350 \end{array} \quad \leftarrow$$

Можно видеть, что в этом случае возникает участок стабилизации частот $(17, 12, 12)$, который обрывается при переходе к знаменю a_{-2} . Ставим левую границу $|$ между a_{-1} и a_{-2} . Теперь, начиная с a_{-1} , делаем новый проход слева направо в поисках правой границы:

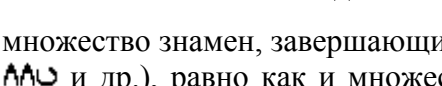
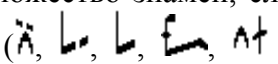
$$\rightarrow \quad \begin{array}{cccc} a_{-1} & a_0 & a_1 & a_2 \\ \begin{array}{c} \text{---} \\ \text{---} \end{array} & \begin{array}{c} \text{---} \\ \text{---} \end{array} & \begin{array}{c} \text{---} \\ \text{---} \end{array} & \begin{array}{c} \text{---} \\ \text{---} \end{array} \\ 435 & 113 & 12 & 12 \end{array} \quad \left| \quad \begin{array}{c} a_3 \\ \begin{array}{c} \text{---} \\ \text{---} \end{array} \\ 5 \end{array}$$

Наблюдаем и в этом случае участок стабилизации частот $(12, 12)$, который обрывается при переходе к a_3 . Ставим правую границу между a_2 и a_3 . Потенциальная семантическая

единица – цепочка  ($F=12$), заключенная между левой и правой границами. Это попевка из семейства “кокиз”.

В общем случае возможны следующие варианты: проставляются обе границы (выделена структурная единица); проставляется лишь одна из них (если правая, то это обычно соответствует попевке с переменным подводом; если левая, то речь чаще всего идет об устойчивой конфигурации, с которой начинаются внутренние разделы песнопений); не выделяется ни одна из границ (структурная единица не обнаружена); наряду с уверенно выделяемыми основными границами наблюдаются более слабо выраженные дополнительные границы (они указывают на наличие более мелких структурных единиц внутри основной).

С помощью описанного подхода хорошо выделяются наиболее массовые попевки гласа, а также устойчивые начальные конфигурации (они короче попевок и не имеют явно выраженной кадансовой структуры). Однако метод дает сбои на *коротких малочастотных* попевках, где отсутствует зона стабилизации частот и теряют смысл понятия “много больше” (\gg), “не намного больше” (\succ) и т.п.

Частичный выход из положения может быть найден, если учесть тот факт, что множество знамен, завершающих попевки, существенно *ограничено* (, равно как и множество знамен, следующих за ними, т.е. начинающих новые структурные единицы ( и т.п.). Если зафиксировать в тексте все вхождения кадансовых (завершающих) знамен, вслед за которыми следуют “начальные” знамена (из второго подмножества), то процедура правостороннего расширения становится ненужной. Имея позиционные привязки в виде кадансовых знамен, упаковываем предшествующие им цепочки, прочитываемые справа налево, в лексикографические деревья со склеенными общими началами (концами попевок). Корнем каждого дерева является одно из кадансовых знамен. Левые концы попевок фиксируются как границы участков стабилизации частот при движении от корня к листьям. Низкочастотные (и, как правило, короткие) цепочки дерева зачастую могут быть идентифицированы *по аналогии* с соседними высокочастотными.

Описанный частотно–позиционный подход существенно менее трудоемок, чем частотный, но также не свободен от недостатков и ограничений. Отметим среди них возможность допущения ошибок при формировании множеств кадансовых и начальных знамен, возможность появления кадансового знамени не в завершающей части попевки (не все из подобных случаев идентифицируются по следующему за кадансовым знамени), а также ориентированность всего подхода лишь на структуры типа “попевка”.

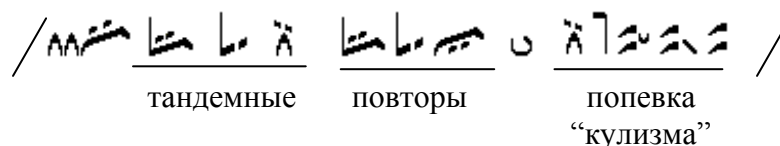
4.5. Максимально длинные повторы в гласе. Самостоятельными структурными единицами обычно являются и максимально длинные фрагменты, *общие* для разных песнопений гласа. Их длины достаточно велики (от 10 до 20 знамен), что говорит о неслучайном характере этих повторов. Часто они являются связующим звеном для песнопений одного жанра. Повторы такого типа представлены в частотных характеристиках высокого порядка ($l \approx l_{\max}$)

Максимальные повторы интересны тем, что представляют собой *конкатенацию* более мелких структурных единиц. Эти единицы могут быть выделены как *общие части* максимальных повторов. Проиллюстрируем это на примере следующей тройки повторов из гласа 2 “Октоиха” (КБ):

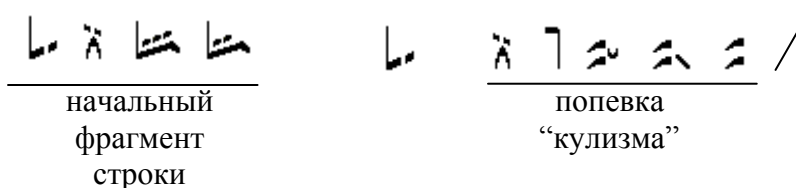
I. ($l=17$)



II. ($l=15$)



III. ($l=11$)



Здесь термин “строка”, многозначный по своей природе, трактуется как “раздел песнопения”, т.е. иначе, чем в п. 4.1. “Косая черта” (/) соответствует паузе в стихотворном тексте, с которой обычно коррелирует музыкальная пауза. Обрамленность повторов (*I*) и (*II*) знаком (/) подчеркивает законченность и самостоятельный статус этих единиц как на уровне стихотворного текста (синтагма), так и в музыкальном отношении (“строка”).

Нетрудно видеть, что повторы *I* и *III* имеют общий фрагмент $S_1 = \text{[musical notation]}$, который обычно фигурирует в начальной части внутренних разделов песнопений (типа “строки”). В свою очередь повторы *II* и *III* имеют общий фрагмент $S_2 = \text{[musical notation]}$, который идентифицируется как попевка “кулизма” [3]. Оба этих фрагмента выделяются как структурные единицы и в соответствии с частотным критерием.

Таким образом, повтор *III* является конкатенацией указанных структурных единиц ($S_1 \text{ [musical notation] } S_2$). Повтор *II* представим в виде $\text{[musical notation]} S_3 S_3 S_2$, где $S_3 S_3$ – тандемно повторяющиеся цепочки $(\text{[musical notation]}) (\text{[musical notation]})$ (знамена $\text{[musical notation]}$ и $\text{[musical notation]}$ в данном контексте имеют одинаковую нотолинейную интерпретацию: восходящее движение двумя четвертными ($|I|=I$)). Повтор *I* может быть записан в виде $S_1 S_4 S_5$, где $S_4 = (\text{[musical notation]})^6$ —серия “стопиц”, соответствующая речитативному участку (простейшая форма тандемного повторения), а $S_5 = \text{[musical notation]}$ идентифицируется в соответствии с [3] как

попевка из семейства “кокиз”. Выделение серии “столиц” в отдельную структурную единицу мотивируется и тем, что следующая за ней попевка встречается в данном гласе и самостоятельно.

Способ выявления структурных единиц по максимальным повторам прост и нагляден. Недостатком его является ограниченность выделяемого с его помощью ассортимента попевок (не все из них входят в состав длинных повторов). Однако характерные начальные фрагменты, набор которых не столь велик, выделяются в достаточно полном объеме, тогда как частотный критерий может на них давать сбои из-за малой длины и нестабильности структурных единиц данного типа.

Заключение.

Рассматривается задача перевода древнерусских церковных песнопений, представленных в знаменной форме записи, в современную нотолинейную. Правила перевода не формализованы, поэтому задача носит дешифровочный характер. Основой для дешифровки служат азбуки “знамя–нота”, подборки попевок, лиц и фит – базовых структурных единиц знаменного распева, а также двознаменники – певческие книги, представленные параллельно в знаменной и нотолинейной форме. Основные затруднения при дешифровке связаны с контекстной зависимостью языка знаменного распева, что обуславливает многозначность соответствия “знамя–нота”. Эта многозначность лишь частично раскрывается в существующих дешифровочных таблицах, которые неполны, порой противоречивы и ориентированы на работу с пометными рукописями (частный случай общей задачи).

Предлагается на основе двознаменников создать компьютерные дешифровочные таблицы нового типа – *L*-граммные азбуки, ориентированные в перспективе на дешифровку беспометных рукописей XVI–XVII веков. Знамена в них представлены со всевозможными лево– и правосторонними контекстами переменной длины. Каждой цепочке знамен сопутствует множество ее нотолинейных интерпретаций с указанием частот их встречаемости в гласе. Расширяя контекст, можно получить однозначно интерпретируемую цепочку.

Разработан, реализован и апробирован на двознаменной певческой книге “Октоих” (XVII в.) алгоритм построения *L*-граммных азбук. Сформулированы критерии отбора наиболее информативных (в плане дешифровки) единиц из *L*-граммных азбук. Набор традиционно используемых единиц (знамен и попевок) предложено пополнить:

1) однозначно интерпретируемыми цепочками знамен, задающими систему звуковысотных ориентиров в каждом гласе, что имеет принципиальное значение при дешифровке беспометных песнопений; 2) тандемными повторами, присутствующими практически в каждом песнопении и несущими в себе элемент тайнозамкнутости; 3) устойчиво повторяющимися начальными фрагментами, не относящимися к категории попевок. При возникновении затруднений с определением типа структурной единицы, в состав которой входит знамя, оно может быть представлено множеством своих контекстов, почерпнутым из *L*-граммных азбук.

ЛИТЕРАТУРА

1. Шабалин Д.С. Певческие азбуки Древней Руси. – Кемерово: Кузбассвуиздат, 1991. – 211 с.
2. Бражников М.В. Древнерусская теория музыки. – Л.: Музыка, 1972. – 423 с.
3. Кручинина А.Н. Попевка в русской музыкальной теории XVII века: Диссертация ...канд. искусствовед. наук: 17.00.02 – Л., 1979.
4. Бражников М.В. Федор Крестьянин. Стихиры. /Публ., расшифровка и исслед. – М., 1974.
5. Зверева С.А. К проблеме расшифровки знаменной нотации конца XVII–XVIII вв. //Проблемы дешифровки древнерусских нотаций. – Л., 1987. – с. 73–89.
6. Шеннон К. Предсказание и энтропия английского печатного текста. //Работы по теории информации и кибернетике. Пер. с англ. /Под редакцией Р.Л.Добрушина и О.Б.Лупанова. – М.: ИЛ, 1963. – с.669–686.
7. Ахо А., Хопкрофт Дж., Ульман Дж. Построение и анализ вычислительных алгоритмов. М., Изд. Мир, 1979, (гл.9).
8. Гусев В.Д., Косарев Ю.Г., Титкова Т.Н. О задаче поиска повторяющихся участков текста //Ассоциативное кодирование. – Новосибирск, 1975. – Вып. 62: Вычислительные системы. – С. 49-71.
9. Гусев В.Д., Титкова Т.Н. Хеширование символьных цепочек в режиме скользящего окна //Анализ последовательностей и таблиц данных. – Новосибирск, 1994. – вып. 150: Вычислительные системы. – С. 94-106.
10. Бахмутова И.В., Гусев В.Д., Титкова Т.Н. Компьютерный анализ древнерусских двознаменников: многозначность соответствий “знамя – нота” и “нота – знамя” //Искусственный интеллект и экспертные системы. – Новосибирск, 1996. – Вып. 157: Вычислительные системы. – С. 68–100.
11. Бахмутова И.В., Гусев В.Д., Титкова Т.Н. Компьютерный анализ древнерусских двознаменников: алфавит знамен и гласоразличительные признаки //Искусственный интеллект и экспертные системы. – Новосибирск, 1997. –Вып. 160: Вычислительные системы. – С. 166–187.
12. И.В.Бахмутова, В.Д.Гусев, Т.Н.Титкова, Б.А.Шиндин. Дешифровочный подход к анализу древнерусских песнопений //Анализ последовательностей и таблиц данных. – Новосибирск, 1994. – вып. 150: Вычислительные системы. – С. 107–130.