



How Expert Users Judge Relevance in Chant Retrieval

A User-Centred Evaluation of SIMSSA's *Cantus Ultimus*

Emma Pope

School of Information Studies, McGill University · Multimodal Interaction Laboratory (MIL) · CIRMMT (Centre for Interdisciplinary Research in Music Media and Technology)

Supervised by Catherine Guastavino

A retrieved match is not yet a relevant result.**What must a symbolic chant retrieval environment make legible for specialized users to compare, verify, and interpret its results?****What evidence do specialists use to decide whether a retrieved relationship is meaningful for their task?**

WHY USERS' RELEVANCE JUDGMENTS MATTER

THE PROBLEM

Retrieval can expose encoded musical relationships through system matching through similarity criteria, but **specialist users assess results through relevance criteria informed by research aims, task context, domain knowledge, and sourcing.**

In chant research, a retrieval environment may support a specialist's larger process of identification, verification, cross-comparison, cataloguing, editorial consultation, or exploration of concordances beyond the location of a single item in a single search.

THE GAP

MIR has developed increasingly sophisticated systems for representing, searching, and providing access to music.

User studies have enriched design and evaluation with evidence about users' needs, behaviours, and judgments.

Medieval and Renaissance chant materials have been extensively encoded, and environments for their retrieval have flourished within symbolic MIR systems development.

The specialist users these retrieval environments were designed for remain relatively undocumented in MIR user studies.

This study examines how these users negotiate relevance through the interpretation and judgment of retrieved chant materials in practice.

(Lee & Cunningham, 2013; Schedl et al., 2014; Weigl & Guastavino, 2013)

STUDY PROPOSITION

For chant specialist users, value may lie both in optimization for topically matched retrieval speed of a source, and in support for cognitively, situationally, and affectively mediated stages of comparison, contextual inspection, verification, uncertainty, and interpretation through multi-field source inspection.

This proposition will be examined empirically. Informed by Laplante & Sauv , 2023; Rousi et al., 2018.

ANALYTIC FRAMEWORK

Task context, confidence in a system, and domain-specific knowledge will be identified through interaction and coding.

Strata of relevance are sensitizing concepts within the framework, not fixed or mutually exclusive coding categories.

DOMAIN-SPECIFIC AND EMERGENT CRITERIA

Potential user group criteria for domain- and task-specific relevance, roughly mapping onto the spheres of **COGNITIVE** and **SITUATIONAL** strata, informed by the system's current fields:

melody · text · liturgical function
notation · manuscript context · provenance
source history · variation · source relationships

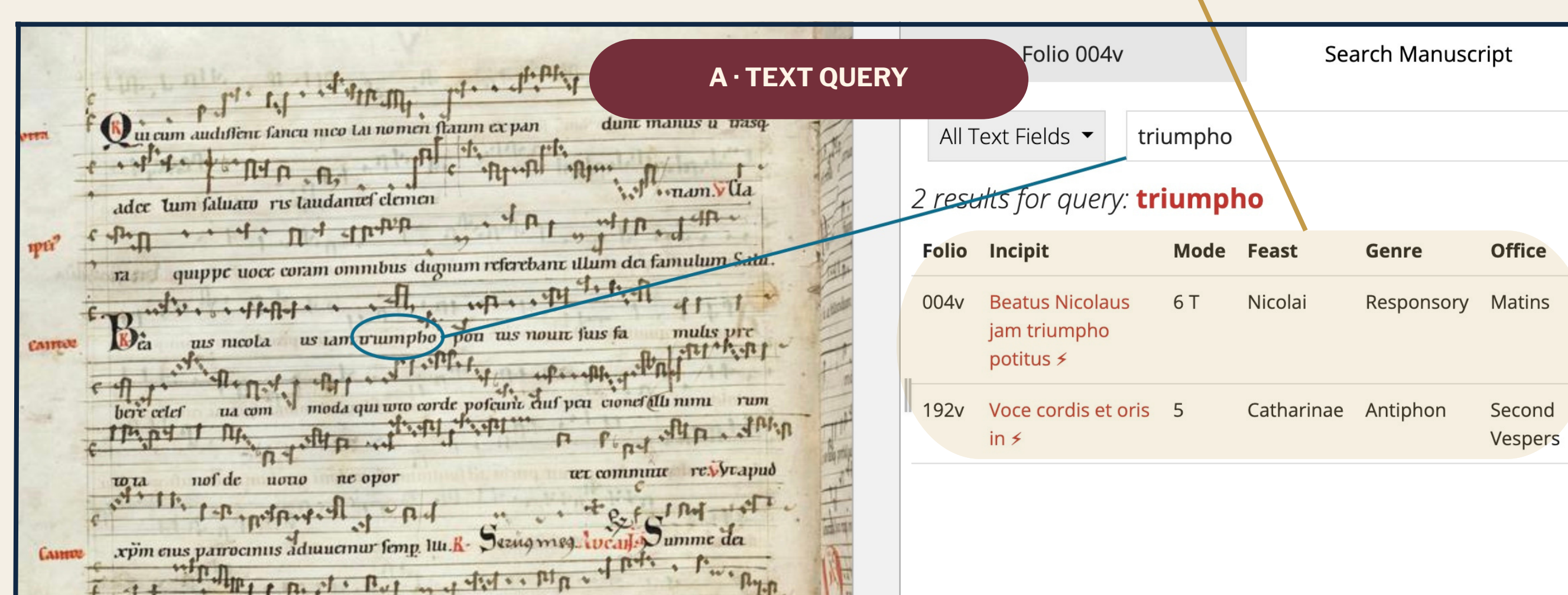
What the Cantus Ultimus interface mediates

as a multi-modal, source-specific environment

Search input and returned metadata include:

text, Volpiano, OMR, Cantus ID, differentiae, mode, feast, genre, office

Text query returning chant records and associated searchable descriptive metadata



Specialist judgment moves across:

image · transcription · record metadata · interpretation

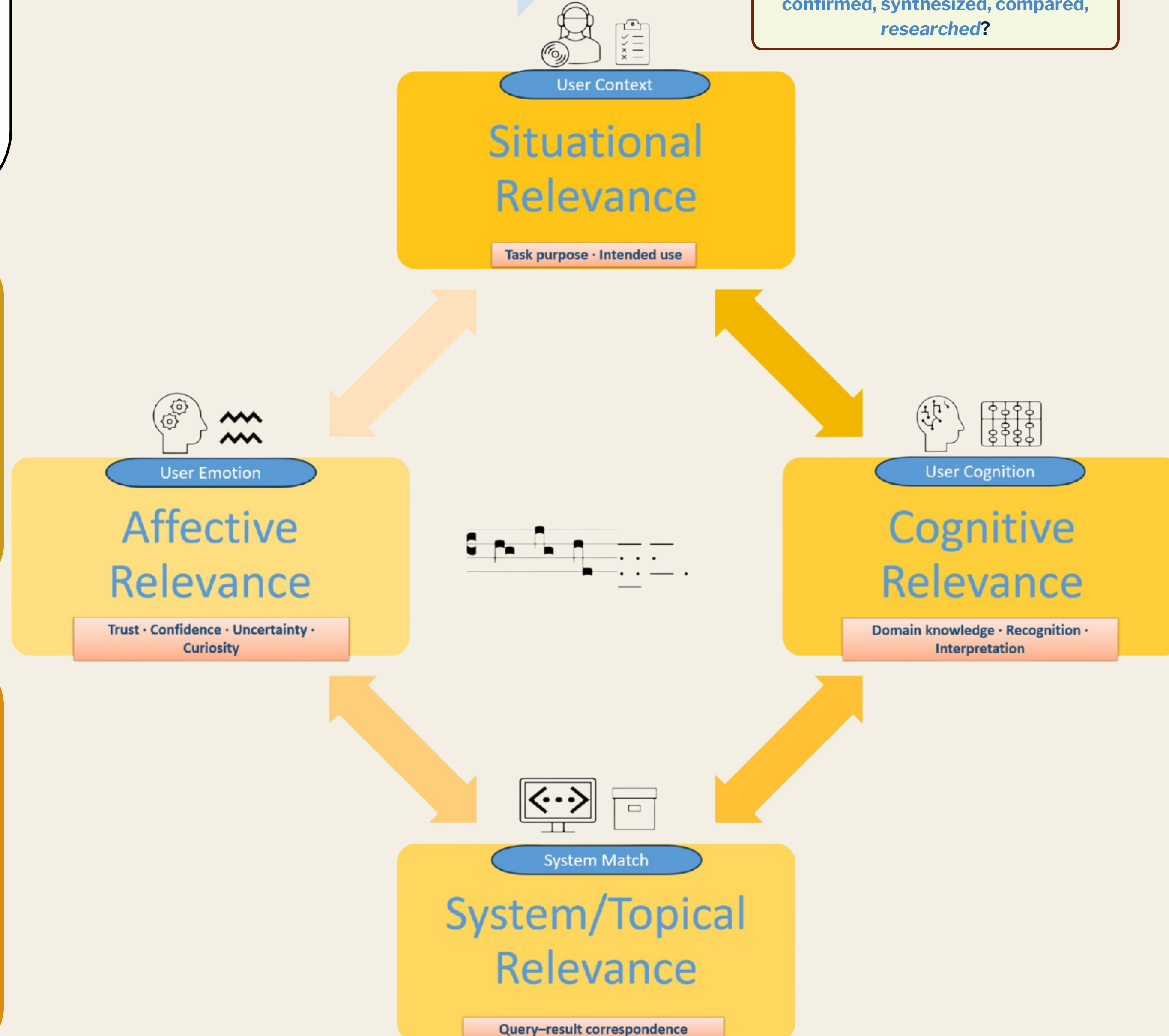
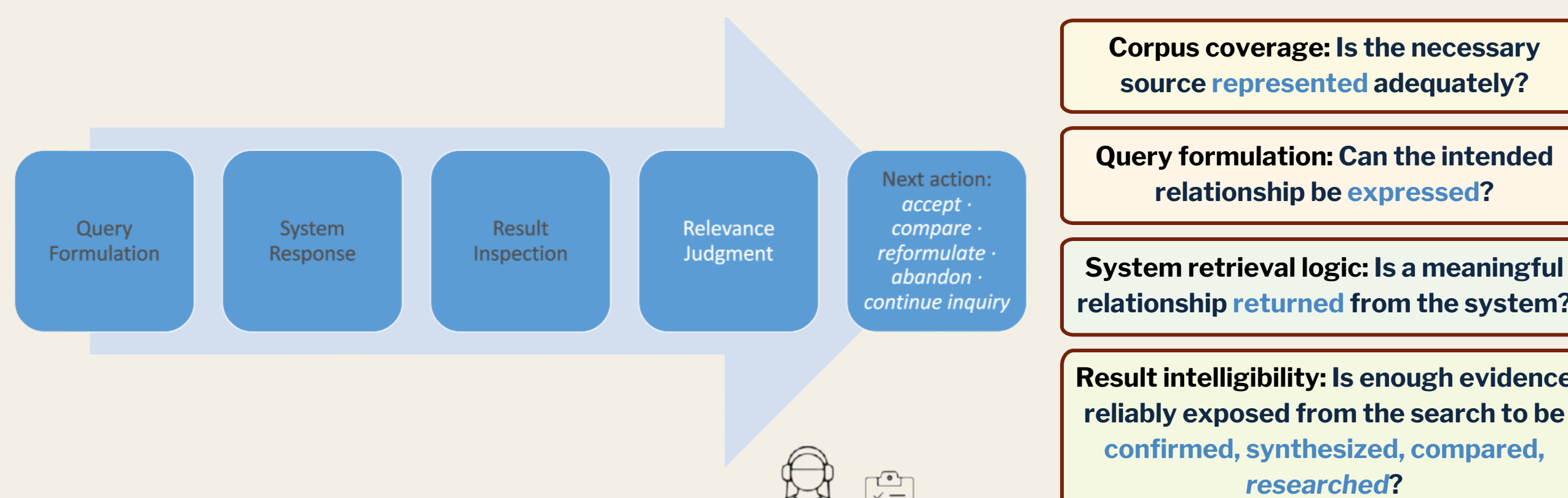
Volpiano (Literal) query returning an instance of a melodic match within a chant record



Source credit: Zwiefalten Antiphoner, Karlsruhe, Badische Landesbibliothek, Musiksammlung, Cod. Aug. perg. LX, fol. 4v.
via Cantus Ultimus through Cantus Database; accessed June 2026.
Illustrative interface views; not study data.

UNIT OF ANALYSIS: THE INTERACTION EPISODE

WHERE CAN INTERACTION BREAK DOWN?



Relevance model adapted from Saracevic's stratified model of relevance (1997, 2007) and its application to MIR by Weigl and Guastavino (2013).

HOW THE STUDY OBSERVES RELEVANCE

VERIFY

Can the participant confirm or contest a proposed identification?*Observe: source evidence · confidence · query-result correspondence*

COMPARE

Can the participant examine variants, concordances, or source relationships?*Observe: comparison strategy · tolerated variation · manuscript context*

EXPLORE

Can the participant identify promising material without a predetermined target?*Observe: orientation · unexpected relevance · next inquiry*

SESSION DESIGN

5-8 specialist users · 45-60 minutes per session

User group defined by familiarity with chant notation, historical sources, chant databases, or related source-based work.

1. Background profile
2. Task-based interaction
3. Concurrent think-aloud
4. Retrospective screen replay
5. Semi-structured interview

DATA GENERATED

Screen action

Concurrent verbalization

Participant expertise

Task context

Researcher observation

Retrospective explanation

ANALYSIS

- Interaction episodes reconstructed across task, query, system response, result inspection, relevance judgment, and next action
- Stratified relevance guides analysis as a sensitizing framework
- Domain-specific and emergent codes remain open

The study evaluates whether users can operate the interface, as well as how they interpret, substantiate, and act upon retrieved relationships, in order to establish where enough evidence is exposed to support confirmation, comparison, and further research.