BEYOND NOTATION: A DIGITAL PLATFORM FOR TRANSCRIBING AND ANALYZING ORAL MELODIC TRADITIONS

Jonathan Myers **Dard Neuman**
Department of Music, University of California, Santa Cruz, Santa Cruz, CA, USA
{JBMyers, DNeuman}@ucsc.edu

ABSTRACT

This paper describes the Interactive Digital Transcription and Analysis Platform (IDTAP): The IDTAP is a web-based application designed to enable users to digitally transcribe, archive, share, and analyze audio recordings of oral melodic traditions. The platform's underlying music-theoretical premises and corresponding data architecture have initially been developed to align with the idiomatic features of Hindustani music (i.e., North Indian classical music). These features necessitate a flexible array of pitch-contour curves; adjustable tuning systems that allow for the representation of a range of microtonal configurations found in both historical and contemporary practice; expressive microtonal pitch inflections between tuning system pitches; and a highly precise rhythmic representation that captures subtle micro-timing nuances, expressive tempo variations, and both metric and non-metric rhythmic structures exactly as performed. The IDTAP's archive, transcription editor, and analysis suite jointly are designed for future expansion to include a range of musical traditions, opening multiple sound collections and archives to digital preservation, pedagogy, and appreciation, as well as statistical, quantitative, and interpretive analysis. The platform and corresponding data architecture equips scholars from a range of disciplinary backgrounds to apply the power of twenty-first-century computational methodologies and large datasets to humanistic and creative endeavors.

1. INTRODUCTION

The IDTAP is an open-source and multi-layered interactive digital archive for oral melodic expressive traditions. [1] The platform allows users to upload audio recordings, transcribe melodic sound intuitively, efficiently, and accurately from those recordings; test the accuracy of their transcription through synthesized audio playback; and analyze the transcriptions qualitatively, quantitatively, and computationally. The broader goal of the IDTAP is to de-center music scholarship and composition from a western theory and notational framework while centering the place of music (or melodic expressive traditions) in humanistic and social scientific studies. Staff notation has long been the primary tool for graphically representing music in academia world-wide, and MIDI remains the dominant data format used in MIR and computational music analysis contexts. Both staff notation and MIDI inherit and reinforce the theoretical premises of Western music theory, particularly the discretization of pitch and rhythm into quantized and equally tempered categories. This limitation makes them ill-suited to accurately represent various aspects of musical information such as intricate glissandi, idiomatic tuning systems, and flexible intonation found in many non-western music traditions.

Since the 1950s, scholars in musicology and ethnomusicology have argued against the use of staff notation for the transcription, composition, and analysis of non-European art music traditions [1–4]. Despite this, many researchers and artists continue to use it when engaging with non-western repertoires. This presents problems that are ethical, empirical, and epistemic. Ethically, staff notation imposes a Eurocentric representation of musical form. Empirically, staff notation flattens out important aspects of many oral-melodic idioms. Epistemically, it discourages scholars without knowledge of staff notation, let alone other methodological tools to include melody and rhythm in their research. This effectively removes from humanistic and social scientific inquiry a central aspect of everyday life—the relationship of organized sound to a range of communities, be they small scale proximate communities or large-scale religious, dynastic, or national imagined communities. These relationships are even more elusive in oral cultures and their secular and spiritual expressive traditions.

By contrast, the IDTAP provides a multi-layered, interactive platform and corresponding store of knowledge, serving as a transdisciplinary research archive that bridges computational media, linguistics, comparative literature, statistics, history, folklore, religious studies, cultural and comparative musicology, ethnomusicology, and music composition. The goal of the IDTAP research team is to expand the platform so that recorded music from a diversity of oral cultures can be both preserved and made available for music-making and research in empirically

[1] The IDTAP is released under the MIT licence. Source code is available at `https://github.com/jon-myers/idtap`. The platform is available at `https://swara.studio`

Attribution: J. Myers and D. Neuman, "Beyond Notation: A Digital Platform for Transcribing and Analyzing Oral Melodic Traditions", in *Proc. of the 26th Int. Society for Music Information Retrieval Conf.,* Daejeon, South Korea, 2025.

analyzable and comparable data formats.

The IDTAP software represents a diversity of melodic contours through our reimagining of a widely held music-theoretical tenet. Rather than treating a fixed pitch note as the atomic unit of musical structure, the IDTAP is organized around a succession of “trajectories”: formally specified archetypal paths between pitches, across a series of pitches, or on a fixed pitch. These trajectories can represent a range of finely calibrated glissandi that appear in many musical traditions outside of the keyboard and staff-oriented Eurocentric musical paradigm.

Trajectories provide a middle-level representation akin to an alphabet of melodic motions. Fundamental frequency data alone, often used directly in analyses, tends to be excessively granular and redundant, offering insufficient analytical utility. Reducing to MIDI or other fixed-pitch representations overly simplifies the musical content, thereby losing critical stylistic nuances. Additionally, trajectories can serve as effective tokens for training AI systems, capturing long-range stylistic continuity and structural coherence more effectively than direct sample-by-sample audio synthesis approaches, which typically struggle to maintain long-range musical consistency [5].

2. RELATED WORK

Recent decades have seen musicology, ethnomusicology, and music information research (MIR) acknowledge the limits of Western staff notation for oral and non-Western traditions. The IDTAP builds on this by offering an interactive platform that combines idiomatic transcription, detailed analytic capability, and corpus-building. By contrast, prior work falls into five categories: (1) automated MIR pipelines focused on feature extraction and one-shot analysis, with no mechanism for building human-informed, reusable corpora; (2) open-access but close-ended datasets and archives; (3) notation or representation frameworks—either idiom-specific or “universal”—that are not integrated into interactive analysis platforms; (4) educational and visualization environments that enrich conceptual understanding but do not let users add recordings or author transcriptions to expand archival corpora; and (5) annotation software that supports audio analysis but lacks idiom-specific transcription or open-access archiving.

2.1 South-Asian MIR

A substantial body of research has explored computational approaches to South-Asian art-music corpora, including research on pitch intonation [6–11], forms of continuous-pitch contours [12–18], and melodic pattern annotation and analysis [19–27]. These studies emphasize automated feature extraction and one-shot analyses. The IDTAP extends this work by enabling manual, idiomatically informed transcription and iterative analysis, supporting the creation of reusable corpora and comparative study.

2.2 South-Asian Archives and Datasets

Music in Motion is a web archive that contains videos that display melodic contour plots with textual annotations of lyrical content scrolling in real time with audio playback [28]. Researchers at Durham University have published a dataset for rhythmic events that is openly available for researchers to use in timing- and rhythm-focused empirical studies [29]. The Music Technology Group (MTG) at Pompeu Fabra University maintains the Saraga collections, the largest annotated open data collections available for computational research on Indian Art Music [30]. Researchers at Georgia Tech have published the Sanidha dataset, the first open-source Carnatic music dataset featuring Studio-quality, multi-track audio recordings with high-definition videos of individual artists’ performances [31]. These archives and resources are, however, close-ended: they do not provide an interactive research environment, as the IDTAP does.

2.3 "Universal" and Non-Western Music Representations

One recent custom notation system which has influenced the design of the IDTAP’s graphical representation is Andrew Killick’s "Global Notation" [32]. Killick’s system offers innovative visual strategies aimed at supporting cross-cultural and comparative music analysis, aspiring towards a global, general framework for representing diverse musical traditions. While the IDTAP has cross cultural applicability—particularly for melismatic-rich traditions—it is not structured for global representation. Rather, it is specifically tailored for idiomatic precision and interactive computational research via software-based implementation.

Similarly, projects encoding Chôngganbo [33], Turkish makam [34], or Carnatic Sargam [35] provide valuable frameworks for notating tradition-specific features, but without integrating those notations into an interactive, analytical software platform as IDTAP does.

2.4 Cross-Cultural Computational Musicology

MTG’s CompMusic [36], Dunya [37], and Musical Bridges [38] have significantly broadened MIR to include diverse world music traditions, offering interactive listening and conceptual exploration tools. IDTAP complements this work by providing interactive, idiom-aware transcription and analysis tools that directly generate reusable, computationally legible data for cross-cultural and comparative study.

2.5 Musicological Analysis Software

Sonic Visualizer, an open-source desktop application, enables researchers to inspect, annotate, and analyze audio through high-resolution waveforms, spectrograms, pitch and rhythm displays, and extensible plugin feature layers [39]. Unlike such general-purpose tools, the IDTAP integrates idiom-specific transcription, structural annotation, and corpus-building in a unified platform.

3. DISCIPLINARY INTERVENTIONS

The fundamental principle of the IDTAP is to diversify music research, representation, creation, criticism, and theory through more expansive transcription forms and corresponding analytics. This principle reflects a commitment to empirical and comparative research that is independent of 19th-century colonial taxonomies, which position European racial and cultural measures as the apex around which other human and cultural creations are assessed and judged. We believe the IDTAP, along with its resultant notations and datasets, will be valuable to a wide range of research/scholarship and teaching/learning.

For one, the IDTAP is designed to diversify humanistic inquiry and support the professionalization of the next generation of researchers. The roles of music transcription and the empirical analysis of music, once distinctive methodological features in the fields of musicology and ethnomusicology, are currently in flux. The musicological turn in U.S. music departments in the 1980s and '90s decentered the role of music theory and other empirical approaches in favor of critical and cultural theoretical approaches [40–43]. This shift occurred partly because music transcription and analysis were reliant on staff notation, even when the music being studied did not align with the European classical tradition for which staff notation was originally developed.

There have been two broad correctives to this situation: the abandonment of transcription and empirical methods altogether, and the creation of custom notation systems. Both solutions, however, raise additional challenges. The turn away from transcription-based research moves the domain of melodic/rhythmic sound analysis away from quantitatively measurable evidence-based research, as well as cutting edge computational technologies based on digital audio research. This shift removes a central part of everyday life from humanistic and social scientific inquiry: the interconnected relationship between melodic traditions and their emotional and affective communities [44]. The second corrective, custom transcription methods, lack commensurate fields of data for quantitative and/or computational analysis, let alone a common language for scholars to engage a broader community of readers and researchers.

The IDTAP is designed, then, as a research tool for scholars of oral melodic traditions and expressive cultures. Statistically oriented researchers can use the data that undergirds this visual representation to ask new, quantitative kinds of research questions. Ethnomusicologists can use "big data" computational approaches to query this archive at a large scale across a range of performances (diachronically or synchronically), or at a small scale through specific measurements within a single performance, or even a single phrase. Music theorists can use the data as a platform for the discovery/construction of new kinds of theoretical knowledge regarding melodic contour, macro- and micro-rhythm, and performer style. Social historians, folklorists, religious studies scholars, and musicologists can supplement their research of (affective) communities, classes, castes, or individuals through the treatment of recorded melodic sound as historical evidence.

The IDTAP also serves as a pedagogical tool for undergraduate and graduate students, as well as independent musicians and music students who want to teach themselves. Whereas generations of practitioners have traditionally learned under the direct tutelage of hereditary or discipular forebears, the IDTAP has the potential to foster a more diverse community of practitioners, including those who weren't born into a musical family and who may not have privileged access to a master-teacher, due to socio-cultural, economic, and/or geographical factors.

Finally, the IDTAP serves as a corrective to the bias currently present in MIR, Music AI research, and corpus studies, which have traditionally focused on Western musical corpora [45]. These epistemological and data-borne biases may have far-reaching consequences over the coming decades, as AI systems play an increasing role in the processes by which music is generated, disseminated, and consumed. To include non-Western and non-notated musical knowledge, their unique forms and structures must first be made computationally legible.

4. IDTAP FRAMEWORK AND THEORETICAL FOUNDATION

The IDTAP has been designed, developed, and implemented with the following components: 1) an archive of recordings, including tools for users to upload, document, and listen to the recordings; 2) an editor interface, which includes a suite of tools for users to compose or transcribe multi-track transcriptions consisting of melodic contours and instrumental, syllabic, and/or textual articulations, as well as to segment and annotate the transcriptions; 3) an archive of transcriptions/notations; 4) data-visualization and data-query functions; and 5) collections organized by user function: research, pedagogy, appreciation, and creative production. The IDTAP user interface and analysis suite enable and make available digital transcriptions for scholars and students who have archival, language, linguistic, musical, or literary research skills—common in Arts, Humanities and Social Science disciplines—but who may not be trained in programming or computational methodologies.

The IDTAP framework is organized around two core concepts: trajectories (see above) and articulations. Articulations coincide with moments of melodic change—such as note onsets, offsets, or instantaneous pitch transitions—and differ according to the instrument being transcribed. Vocal articulations use International Phonetic Alphabet (IPA) [46] symbols, allowing users to annotate each trajectory with a starting consonant, vowel, and ending consonant, while instrumental articulations distinguish playing techniques such as plucking, bowing, blowing, pressing, or hammering. Each instrument, vocal or non-vocal, has its own set of available trajectories and articulations, reflecting its unique expressive characteristics.

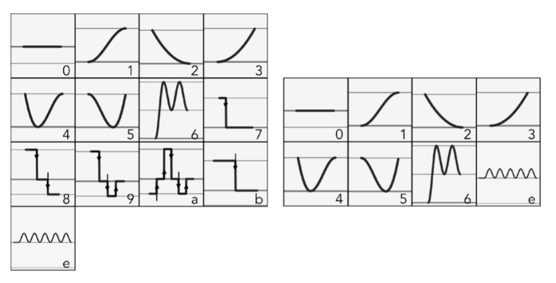

Figure 1. The thirteen trajectories used in *sitar* (plucked string) transcriptions and the eight trajectories used in vocal and *sarangi* (bowed string) transcriptions.

5. IDTAP EDITOR INTERFACE

The "Editor" interface allows users to transcribe a specific recording—or make notations from scratch—by inserting trajectories onto a two-dimensional time-by-pitch graph. To transcribe, users place orientation points on the graph with an overlaid spectrogram and/or pitch trace and select from an alphabet of trajectories—see Figure 1—to find the curve that most closely fits the melodic contour. For certain trajectories, the slope of the curve can be adjusted to better match the melodic contour being transcribed. Orientation points can also be "offset" to represent microtones (*shrutis*)—or spaces between scale tones—and melodic contours that move between scale tones and microtones (*andolans*). Figure 2 illustrates two transcriptions of a section of the Hindustani *thumri* composition, "Babul Mora," performed by Begum Akhtar. As shown, the IDTAP is highly attuned to the nuances of movements both between and on scale tones.

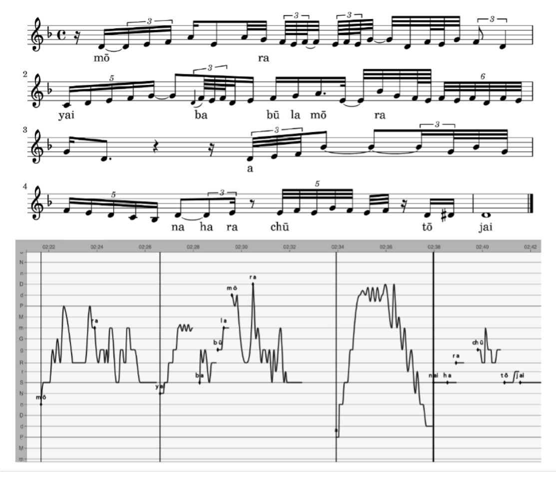

Figure 2. Transcription of the same selection of the song, *Babul Mora*, using staff and IDTAP notations.

5.1 Graphical Display

The IDTAP editor interface represents pitch on a graph using a logarithmic frequency scale with adjustable upper and lower bounds. Users can customize the visual elements extensively, including changing colors for clarity and accommodation. Horizontal lines indicate the primary scale degrees (*swara*) of the current scale system (*raga*), and their precise microtonal tuning can be adjusted to match the specific pitch configurations of individual performances or traditions.

The graphical display supports up to four instrumental or vocal tracks simultaneously, each presented in distinct user-selected colors. Various instrumental articulations are represented through symbolic notation: for *sitar*, standard plucks are displayed as triangles, *chikari* string plucks as "X" symbols, and string mutes as vertical lines at the endpoints of trajectories; for vocal tracks, consonants appear as diamond shapes.

Additionally, users can toggle visual overlays, including a spectrogram [47] and a pitch trace [48]. The spectrogram visualization is adjustable in both intensity and colormap, enhancing its analytical utility. These graphical tools enable users to visually verify transcription accuracy, clearly discern nuanced melodic behaviors, and effectively compare multiple performance layers.

5.2 Meter

The IDTAP interface supports flexible specification, visualization, and adjustment of hierarchically nested metrical structures (e.g. meter or *taal*). Users can define metrical templates and place them precisely at desired moments within a transcription. The platform allows fine-tuning of micro-rhythmic sub-division placement to align closely with the natural temporal fluctuations (rubato) inherent in performance practice. Additionally, users can adjust tempo dynamically, set varying subdivisions within each hierarchical layer, and visually insert alignment points directly onto the transcription. An integrated algorithm automatically calculates the optimal metrical alignment, determining the appropriate tempo and beat placement based on these user-defined reference points.

5.3 Tuning

The IDTAP interface includes flexible tuning tools that accommodate the precise microtonal requirements of diverse musical traditions. Users can globally adjust the fundamental frequency of the transcription's tonic pitch ("sa"), automatically shifting the vertical alignment of associated visual elements, including horizontal pitch-reference lines and spectrogram boundaries. Additionally, each scale degree can be individually fine-tuned, enabling precise pitch adjustments in cents relative to equal temperament. To facilitate accurate tuning, sine-tone playback for each pitch can be activated, allowing users to aurally align these reference tones with the original audio recording. Simultaneously, visual alignment between adjusted pitch-reference lines and corresponding melodic contours on the spectrogram supports users in refining tuning configurations.

5.4 Segmentation and Annotation

The IDTAP interface provides robust tools for segmenting transcriptions at multiple hierarchical levels, enabling precise structural representation and detailed analytical annotation. Transcriptions can be divided vertically into two distinct hierarchical layers: Sections and Phrases. Section divisions represent broader structural elements (e.g., compositions or improvisational segments), with thick vertical lines, whereas Phrase divisions represent finer subdivisions within Sections, marked by thinner vertical lines. Users can insert and adjust these divisions intuitively, using simple mouse actions or keyboard shortcuts.

Annotations for each segment or phrase can be applied using the Label Editor, allowing for detailed descriptive analysis. For Sections, users may label broader categories such as 'Improvisation,' 'Composition,' '*Alap*,' or specify a particular type of composition. Phrase-level annotation offers even greater granularity, enabling users to select from multiple descriptors across several dimensions: Phrase-Types (compositional forms); Elaboration-Types (improvisational forms); Articulation-Types (e.g., if a vocalist is singing lyrics (*bols*)), Note-Names (*sargam*), Non-Lexical syllables (*aakar*, *nom-tom*); and Incidentals (e.g., tuning, talk, or conversations by musicians). Multiple labels can be concurrently applied to any single phrase, reflecting the complexity of performance practice.

5.5 Synthesis

The IDTAP integrates a real-time synthesis module to facilitate transcription. Implemented with the Web Audio API, this module renders the current transcription as audio, allowing users to audition the synthesized rendition alongside the original recording in order to iteratively refine a transcription's accuracy.

For vocal transcriptions, the IDTAP employs the Klatt Synthesizer, a formant synthesis model designed to simulate human speech and vocal timbres [49]. Plucked-string instruments, such as the *sitar*, are synthesized using the Karplus-Strong algorithm, a physical modeling technique effective for simulating plucked string acoustics [50]. Bowed string instruments, such as the *sarangi*, are synthesized using a custom-designed engine that incorporates dynamically modulated feedback loops, resonant filtering, and noise-based excitation to emulate acoustic interactions typical of bowed chordophones.

In transcriptions involving multiple instruments, each instrument is assigned a dedicated synthesizer. A mixer interface enables users to independently adjust the volume levels of the original recording and synthesized tracks, facilitating precise auditory comparisons suited for detailed transcription and analysis tasks. Additional features include looping of specific transcription segments, adjustable playback speeds, and pitch shifting capabilities that alter the fundamental frequency of both synthesized audio and original recordings. These functionalities support transcription workflows that require alignment with instruments tuned to different reference tonics.

6. ANALYSIS, DATA VISUALIZATION, AND QUERY PROCESS

The "Analysis" interface includes interactive tools for users to engage with transcription data heuristically. These tools include: 1) a query system that isolates and displays specific sequences from a transcription based on an unlimited number of filtering criteria; 2) a pitch occurrence data-visualization tool that displays the relative proportion of time spent on particular pitches, and segmented by section, phrase, or durational window; 3) a "pitch pattern" visualization tool that displays the relative frequencies of specific pitch patterns throughout a transcription. This interface supports both emic and etic analyses: Emic approaches allow analysts to annotate and analyze based on "insider" sensibilities, categories, and concepts, while etic approaches enable analysis based on "neutral" parameters [51].

6.1 Pitch Prevalence

The pitch prevalence interface allows users to generate visualizations representing the prevalence of particular pitches over the course of a transcription. This is achieved by splitting the transcription into durational windows according to one of three user-selected segmentation schemes. One segmentation scheme—by duration—is etic, as it divides the transcription into windows of user-defined duration. This method aims to overcome various biases: both insider knowledge of traditional conceptions of phrase structure as well as outsider intuitive perceptions of phrase structure. Two of the segmentation schemes—by section and by phrase—are emic approaches, relying on an analyst's insider knowledge and their decisions on where to place analytical intentions in explicitly entering phrase and section divisions throughout the process of transcription. For the emic segmentation approaches, the user is also able to filter the segments to include or exclude particular section or phrase types, as annotated by the user during transcription. With the section segmentation, a user may choose to include or exclude section types: in the case of Hindustani music, different forms of composition and improvisation.

For example, Figure 3 displays the pitch prevalence and duration of the first six sections and eleven phrases of the *thumri* song, "Babul Mora," performed by Begum Akhtar, along with an annotated description of each section and phrase type, allowing users to see at a glance both the formal structure and pitch distribution of the performance.

In addition to the type of segmentation, users can toggle between: a) a visualization to tabulate the prevalence of each pitch chroma (i.e. pitch class, the quality of a pitch as abstracted from its octave registration) instead of individual pitches; b) a vertically condensed display, which excludes empty vertical space for pitches not included in the transcription; c) a "heatmap" grayscale gradient that represents the relative prevalence of elements, rather than showing the most common (mode) element in dark gray and the rest in light gray.

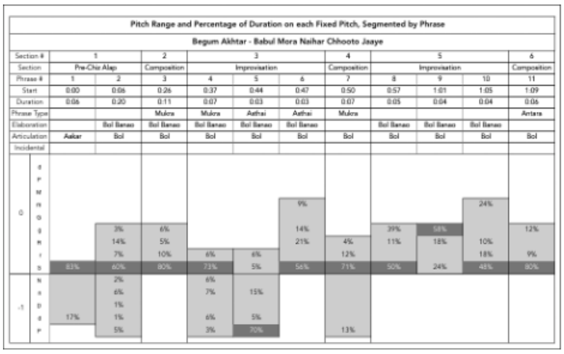

Figure 3. Pitch Prevalence Data Visualization.

6.2 Query Display

The query display interface allows users to create multi-part queries to filter a transcription, displaying only those segments that match specific criteria. For example, Figure 4 demonstrates the query function and output for every instance when Begum Akhtar sings “babul” (father). The query, in this case, looks for a phrase containing a starting consonant “ba,” a vowel “u,” and an ending consonant “l”. This type of query could be used for pedagogical or analytical investigations into how an artist expresses a composition—where word and melody are fixed—in multiple ways, visually examining the interplay between composition and improvisation.

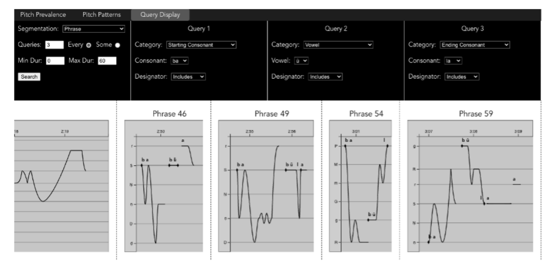

Figure 4. Query Functions and display example.

Users can construct queries based on a wide range of categories, including: Pitch, Strict Pitch Sequence (pitches must appear in this exact order without any gaps between them), Loose Pitch Sequence (pitches must appear in this order, but may have gaps between them), Trajectory, Strict Trajectory Sequence, Loose Trajectory Sequence, Section Type, *Alap* Type, Composition Type, Tempo, *Tala* (descriptor of metric structure and hierarchy, akin to a Time Signature in western music), Phrase Type, Elaboration Type, Incidental Type, or Articulation Type.

For each category, users can select a designator: “includes,” “excludes,” and, if applicable, “starts with,” or “ends with.” For example, a query might include all phrases that contain the pitch “Sa” or exclude all phrases that contain a specific trajectory, such as a “*krintin*-slide.”

This querying system is especially powerful when users combine multiple queries, allowing them to focus on specific analytical questions or hypotheses. Users can choose to display those segments that satisfy all of the queries, or just at least one of the queries. Once the desired phrases are displayed, analysts can explore the filtered results and draw insights from their observations.

6.3 Pitch Patterns

The Pitch Patterns interface enables users to select a specific type of segmentation and pattern size. It then counts and records the number of occurrences of each unique pitch sequence within each segment, based on the chosen pattern size, and displays this information alongside a graphical depiction of the contours of each sequence. This kind of display allows analysts to intuitively search for common patterns of various sizes in different segments throughout a performance, thus revealing some essence of the style of the melodic elaboration.

6.4 Analysis Suite Implications

Together, the types of analysis enabled by these queries, data visualizations, and datasets are multiple. To name just a few, users can examine a particular recording, section from a recording, or body of recordings associated with an individual or group of musicians from a range of traditions. Research topics could include: relationships between lyric-text and melody, or syllable and melodic movement; empirically measurable changes of a musician’s performance style over time: throughout an individual performance or across many performances; analysis across instruments, venues, styles, performance sections, time-periods, geographic areas, recording media, or other clearly specified method of slicing a corpus. These three analysis tools will be enhanced and added to as our project team continues to engage with and respond to the needs of our research colleagues. The IDTAP also lets users export data in JSON or CSV, so they can conduct quantitative analyses in whichever software or programming environment they prefer.

7. CONCLUSION

Existing musical representation schemes inevitably reflect their socio-historical contexts, embedding assumptions about musical structures, forms, and meanings. The IDTAP provides a new framework to represent and analyze oral melodic traditions precisely and flexibly, responding directly to the nuanced demands of Hindustani music. By centering expressive musical features typically overlooked by traditional notation systems, the platform aims to foster new research methodologies, pedagogical opportunities, and creative practices. As computational approaches to music scholarship grow increasingly significant, platforms like IDTAP help ensure that diverse musical traditions remain visible, analyzable, and creatively vital in the digital age.

8. ACKNOWLEDGMENTS

This work was supported by the National Endowment for the Humanities (grant HAA-290356-23) and by several programs at the University of California, Santa Cruz: the Arts Research Institute, the Office of Research, and the Hasan and Ali Akbar Khan Endowment for Classical Indian Music. We thank the ISMIR reviewers for providing valuable feedback on this research. We dedicate this paper to the memory of our colleague, mentor, and friend Larry Polansky.

9. REFERENCES

[1] C. Seeger, "Toward a Universal Music Sound-Writing for Musicology," *Journal of the International Folk Music Council*, vol. 9, pp. 63–66, Jan. 1957.

[2] N. M. England et al., "Symposium on Transcription and Analysis: A Hukwe Song with Musical Bow," *Ethnomusicology*, vol. 8, no. 3, pp. 223–233, Sep. 1964.

[3] M. Hood, *The Ethnomusicologist*. New York, 1971.

[4] T. Ellingson, "Transcription," in *Ethnomusicology: An Introduction*, H. Myers, Ed. Basingstoke, 1992, pp. 153–164.

[5] K. Goel et al., "It's Raw! Audio Generation with State-Space Models," *International Conference on Machine Learning, PMLR*, 2022.

[6] S. Şenturk et al., "A Score-Informed Computational Description of Svaras Using a Statistical Model," in *Proceedings of the 13th Sound and Music Computing Conference*, Hamburg, Germany, 2016.

[7] G. k. Koduri et al., "Characterization of Intonation in Carnatic Music by Parametrizing Pitch Histograms," in *ISMIR*, Porto, Portugal, 2012.

[8] J. Serrà et al., "Assessing the Tuning of Sung Indian Classical Music," in *ISMIR*, 2011.

[9] M. M. Komaragiri, "An Empirical Study of Intonation in Raga Kalyani," in *Proceedings of the International Symposium "IT-to-Mass: Speech and Music"*, Bhubaneswar, India, 2005.

[10] K. Akant et al., "Measuring Frequencies of Shrutis in Indian Classical Music," in *Proceedings of the 9th International Conference on Emerging Trends in Engineering and Technology—Signal and Information PRocessing*, 2019.

[11] M. M. Komaragiri et al., *Pitch Analysis in South Indian Music: With a Critical Examination of the Theory of 22 Śruti-s*. New Delhi: Munshiram Manoharlal, 2013.

[12] V. S. Viraraghavan, R. Aravind, and H. A. Murthy, "A Statistical Analysis of Gamakas in Carnatic Music," in *Proc. ISMIR*, Suzhou, China, Oct. 2017, pp. 243–249.

[13] R. Morris, "Tana Varnam-s: An Entry into Rāga Delineation in Carnatic Music," *Analytical Approaches to World Music*, vol. 1, no. 1, pp. 1–27, 2011.

[14] T. Nuttall, X. Serra, and L. Pearson, "Svara-Forms and Coarticulation in Carnatic Music: An Investigation Using Deep Clustering," in *Proceedings of the 11th International Conference on Digital Libraries for Musicology (DLfM)*, Stellenbosch, South Africa, Jun. 2024.

[15] T. Nuttall, X. Serra, and L. Pearson, "Svara-Forms in Carnatic Music: Contextual Influences on the Performance of Svara," in *Workshop on Indian Music Analysis and Generative Applications (WIMAGA)*, satellite workshop of IEEE ICASSP, Hyderabad, India, Apr. 2025.

[16] L. Pearson, "Coarticulation and Gesture: An Analysis of Melodic Movement in South Indian Raga Performance," *Music Analysis*, vol. 35, no. 3, pp. 280–313, 2016.

[17] M. Subramanian, "Carnatic Rāgam Thodi – Pitch Analysis of Notes and Gamakams," *Journal of the Sangeet Natak Akademi*, vol. 41, no. 1, pp. 3–28, 2007.

[18] M. Subramanian, "Carnatic Music – Automatic Computer Synthesis of Gamakams," *Journal of the Sangeet Natak Akademi*, vol. 43, no. 3, pp. 28–36, 2009.

[19] T. Nuttall et al., "In Search of Sañcaras: Tradition-Informed Repeated Melodic Pattern Recognition in Carnatic Music," in *Proceedings of the 23rd International Society for Music Information Retrieval Conference*, Bengaluru, India, 2022.

[20] K. K. Ganguli et al., "Data-Driven Exploration of Melodic Structure in Hindustani Music," in *Proceedings of the 17th International Society for Music Information Retrieval Conference*, New York, NY, USA, 2016, pp. 605–611.

[21] S. Gulati, *Computational Approaches for Melodic Description in Indian Art Music Corpora*. Ph.D. dissertation, Universitat Pompeu Fabra, Barcelona, Spain, 2017.

[22] S. Gulati et al., "Mining Melodic Patterns in Large Audio Collections of Indian Art Music," in *Proceedings of the International Conference on Signal Image Technology and Internet Based Systems*, Marrakech, Morocco, 2014.

[23] S. Gulati et al., "Discovering Raga Motifs by Characterizing Communities in Networks of Melodic Patterns," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Shanghai, China, 2016, pp. 286–290.

[24] V. Ishwar et al., "Motif Spotting in an Alapana in Carnatic Music," in *Proceedings of the 14th International Society for Music Information Retrieval Conference*, Curitiba, Brazil, 2013, pp. 499–504.

[25] T. Nuttall et al., "The Matrix Profile for Motif Discovery in Audio – An Example Application in Carnatic Music," in *Proceedings of the 15th International Symposium on Computer Music Multidisciplinary Research*, Tokyo, Japan, 2022.

[26] P. Rao et al., "Melodic Motivic Analysis of Indian Music," *Journal of New Music Research*, vol. 43, no. 1, pp. 115–131, 2014.

[27] P. Rao et al., "Classification of Melodic Motifs in Raga Music with Time-Series Matching," *Journal of New Music Research*, vol. 43, no. 1, pp. 115–131, 2014.

[28] S. Rao and W. van der Meer, "Music in Motion: The automated transcription for Indian music (AUTRIM) project by NCPA and UvA," 2017. [Online]. Available: `https://autrimncpa.wordpress.com/`. [Accessed Jan. 16, 2024].

[29] M. Clayton, L. Leante, and S. Tarsitani, "IEMP North Indian Raga," *Database of annotated audiovisual recordings*, 2018. [Online]. Available: `https://doi.org/10.17605/OSF.IO/KS325`.

[30] A. Srinivasamurthy et al., "Saraga: Open Datasets for Research on Indian Art Music," *Empirical Musicology Review*, vol. 16, no. 1, pp. 85-98, 2021.

[31] V. V. Krishnan et al., "Sanidha: A Studio Quality Multi-Modal Dataset for Carnatic Music," in *ISMIR*, 2024.

[32] A. Killick, "Global Notation as a Tool for Cross-Cultural and Comparative Music Analysis" *Analytical Approaches to World Music*, vol. 8, no. 2, pp. 235–279, 2020.

[33] J. Lee, J. Downie, and A. Renear, "Representing Korean Traditional Musical Notation in XML," in *ISMIR*, 2002.

[34] M. Karaosmanoğlu, "A Turkish Makam Music Symbolic Database for Music Information retreival: SymbTr," in *ISMIR*, 2012.

[35] S. Mammen, et al., "iSargam: Music Notation Representation for Indian Carnatic Music," EURASIP Journal on Audio, Speech, and Music Processing, vol. 2016, no. 5, 2016.

[36] X. Serra, "A Multicultural Approach in Music Information Research," in *ISMIR*, pp. 151-156, 2011.

[37] A. Porter, M. Sordo, and X. Serra, "Dunya: A System for Browsing Audio Music Collections Exploiting Cultural Context," in *ISMIR*, 2013.

[38] Universitat Pompeu Fabra, Music Technology Group, "Musical Bridges – Bridging cultures through music understanding," 2021. [Online]. Available: `https://www.upf.edu/en/web/musicalbridges/`. [Accessed: 13-Jun-2025].

[39] C. Cannam, C. Landone, and M. Sandler, "Sonic Visualiser: An Open Source Application for Viewing, Analysing, and Annotating Music Audio Files," in *Proceedings of the ACM Multimedia International Conference*, 2010.

[40] S. Mclary, *Feminine Endings: Music, Gender, and Sexuality*. Minneapolis: Univ. Minn. Press, 1991.

[41] J. Kerman, *Contemplating Music: Challenges to Musicology*. Cambridge, MA: Harvard University Press, 1985.

[42] L. Goehr, *The Imaginary Museum of Musical Works: An Essay in the Philosophy of Music*. Oxford: Clarendon Press, 1992.

[43] M. Citron, *Gender and the Musical Canon*. Cambridge: Cambridge University Press, 1993.

[44] L. Gandhi, *Affective Communities: Anticolonial Thought, Fin-de-siècle Radicalism, and the Politics of Friendship*. Durham: Duke University Press, 2006.

[45] P. E. Savage, "An overview of cross-cultural music corpus studies," in *Oxford Handbook of Music and Corpus Studies*, D. Shanahan, A. Burgoyne, and I. Quinn, Eds. New York: Oxford University Press, [In press]. [Online]. Available: `http://doi.org/10.31235/osf.io/nxtbg`.

[46] International Phonetic Association, *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press, 1999.

[47] C. Schörkhuber, A. Klapuri, N. Holighaus, and M. Dörfler, "A MATLAB toolbox for efficient perfect reconstruction time-frequency transforms with log-frequency resolution," in *Proc. 53rd International Conference: Semantic Audio*, Audio Engineering Society, Jan. 2014.

[48] J. Salamon and E. Gómez, "Melody extraction from polyphonic music signals using pitch contour characteristics," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 6, pp. 1759–1770, 2012.

[49] D. H. Klatt, "Software for a cascade/parallel formant synthesizer," *Journal of the Acoustical Society of America*, vol. 67, no. 3, pp. 971–995, 1980.

[50] K. Karplus and A. Strong, "Digital synthesis of plucked-string and drum timbres," *Computer Music Journal*, vol. 7, no. 2, pp. 43–55, 1983.

[51] M. Harris, "History and Significance of the Emic/Etic Distinction," *Annual Review of Anthropology*, vol. 5, pp. 329–350, 1976.