

## Decoding the Song: Histogram-Based Paradigmatic and Syntagmatic Analysis of Melodic Formulae in Hungarian Laments, Torah Trope, 9<sup>th</sup> Century Plainchant and Koran Recitation

Dániel Péter Biró<sup>1</sup>, Steven Ness<sup>2</sup>, Matthew Wright<sup>1,2</sup>,  
W. Andrew Schloss<sup>1</sup>, George Tzanetakis<sup>2</sup>

University of Victoria, School of Music<sup>1</sup> and Department of Computer Science<sup>2</sup>  
Engineering/Computer Science Building (ECS), Room 504,  
PO Box 3055, STN CSC, Victoria, BC, Canada V8W 3P6

### Abstract

The development of musical notation and the changing relationship between textual syntax and musical semiotics were inherently connected to the transformation of a culture based on oral transmission and ritual to one based on writing and hermeneutic interpretation. Along this historical continuum, notation functioned either to reconstruct a previous, remembered melody or to construct a newly composed melody. For the chant scholar the question arises as to when and under what conditions melodic formulae became solidified as musical material. In the present study we examine examples from improvised, partially improvised, partially notated and gesture-based notational chant traditions: Ashkenazi Torah cantillation, Ninth Century St. Gallen plainchant, and Koran recitation. We explore examples from these various traditions through computational tools for paradigmatic analysis of melodic formulae and gesture.

Exploring the functionality of melodic gesture, musical syntax and musical semiotics in the specific contexts of speaking, singing, reading and writing enhances the comprehension of the relationship between melodic formula and textual syntax within these divergent forms of religious chant.

### Formula, Gesture and Syntax

These various types of chant employ *melodic formulae*, figures that define certain melodic identities that help to define syntax, pronunciation and expression. Each tradition, and the resulting melodic framework thereof, is governed by the particular religious context for performance.

Jewish Torah trope is “read” using the twenty-two cantillation signs of the *te’amei hamikra*.<sup>1</sup>

<sup>1</sup> The term “*te’amei hamikra*” means literally “the meaning of the reading.”

developed by the Masorite rabbis<sup>2</sup> between the sixth to the ninth centuries. The melodic formulae of Torah trope govern syntax, pronunciation and meaning and their clearly identifiable melodic design, determined by their larger musical environment, is produced in a cultural realm that combines melodic improvisation with fixed melodic reproduction within a static system of notation.

The rules for Koran recitation are not determined by text or by notation but by the act of recitation itself: <sup>3</sup> here the hierarchy of spoken syntax, expression and pronunciation play a major role in determining the vocal styles of

<sup>2</sup> “Originally, the biblical books were written as continuous strings of letters, without breaks between words. This led to great confusion in the understanding of the text. To ensure the accuracy of the text, there arose a number of scholars known as the Masoretes in the sixth century CE, and continuing into the tenth century.” (Widgoder et al 1989).

<sup>3</sup> Heidi Zimmermann, *Tora und Shira: Untersuchungen zur Musikauffassung des rabbinischen Judentums* (Bern: Peter Lang, 2000) 128, “Die primäre Funktion des Korans als Rezitationstext – darauf verweist schon der Name – macht seine Mittelstellung zwischen rein schriftlicher und rein mündlicher Literatur deutlich.” “The primary function of the Koran as a recited text – its very name refers to this function – clearly presents its place in between written and purely oral literature.” (English translation by Biró) “Im Unterschied zur Hebräischen Bibel gilt der Koran jedoch also in Sprache und nicht als in Schrift offenbart. Er ist zwar unmittelbar nach dem Tod des Propheten Muhammad verschriftet worden und in Buchform überliefert, wird aber erst sekundär als *al-kitāb* “das Buch” verzeichnet.” Der primäre Name ‘koran’ ist – wie das hebräische *miqra* – abgeleitet aus der Wurzel q-r-: ‘lesen’, impliziert aber hier nicht die visuelle Aufnahme von Text. Vielmehr drücken sich darin Konzepte wie ‘aussprechen, ausrufen, rezitieren’ aus, so dass eine adäquate Übersetzung für Koran (Qur’ ān) ‘das zu Rezitierende’ lauten könnte.” “Unlike the Hebrew Bible, the Koran is considered to be revealed as recited speech and not as text. Although the Koran was written just after the death of Mohamed and later transmitted in book form, the name *al-kitāb* “the book” is used only as a secondary title. Like the Hebrew *miqra* the primary name ‘Koran’ derives from the root q-r-, i.e., “reading”: the visual implication of text is not implied with this root. Rather the concepts “pronounce, calling, reciting” are expressed with the word, so that an adequate translation of Koran (Qur’ ān) could be “the recited.” (English translation by Biró)

*Tajwīd*<sup>4</sup> and *Tartīl*.<sup>5</sup> The resulting melodic phrases, performed not as “song” but “recitation” are, like those of Jewish Torah trope, determined by both the religious and larger musical cultural contexts.

Textual hierarchy can also be observed in the early plainchant neumes, which evolved from a logogenic culture that was based on textual memorization<sup>6</sup> (Treitler 1982) within this culture the memorized singing of chants was central to the preservation of a tradition that developed over centuries.<sup>7</sup> Already in the ninth century the technology of writing was advanced enough to allow for new degrees of textual nuance. Here the ability for formulae to transcend textual syntax is at hand, pointing to the possibility for melodic autonomy from text.

Chant scholars have investigated historical and phenomenological aspects of chant formulas to discover how improvised melodies might have developed to become stable melodic entities, paving the way for the development of

notation.<sup>8</sup> A main aspect of such investigations has been to

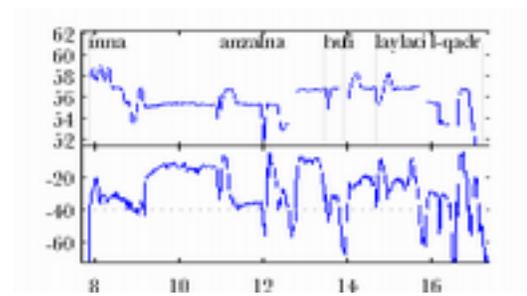


FIGURE 1. Pitch and Energy Contours

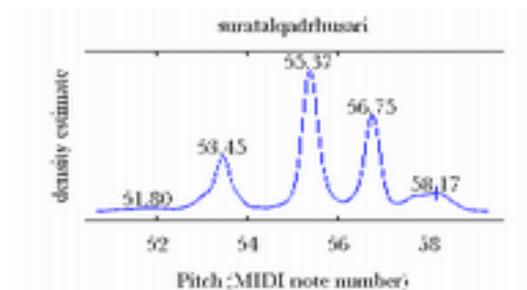


FIGURE 2. Recording specific scale derivation

explore the ways in which melodic contour defines melodic identities (Karp 1998). We hope that our computational tools will allow for new possibilities for paradigmatic and syntagmatic chant analysis in both culturally defined and cross-cultural contexts. This might give us a better sense of the role of melodic gesture in melodic formulae and possibly a new understanding of the evolution from improvised to notation-based singing in and amongst these divergent chant traditions.

<sup>4</sup> “Tajwīd, the system of rules regulating the correct oral rendition of the Qur’an, governs many parameters of sound production. These include precise duration of syllable, vocal timbre and pronunciation, with characteristic use of nasality and special techniques of vibration. Echoing silences between text sections add to the dynamic nature of presentation. Public Qur’anic recitation has a distinctive sound which has been profoundly influential as an aesthetic ideal.” Eckhard Neubauer, Veronica Doublday: ‘Qur’anic recitation,’ Grove Music Online ed. L. Macy. (Accessed 6 June 2008) <<http://www.grovemusic.com.ezproxy.library.uvic.ca>>

<sup>5</sup> “Tartīl, another term for recitation, especially implies slow deliberate attention to meaning, for contemplation.” Eckhard Neubauer, Veronica Doublday: ‘Qur’anic recitation,’ Grove Music Online ed. L. Macy. (Accessed 6 June 2008) <http://www.grovemusic.com.ezproxy.library.uvic.ca>

“As explained by Sayx al-Sahhar, tartīl has some melody but not “complete voice.” The tempo is relaxed but not slow.” (Nelson, 20)

<sup>6</sup> Peter Jeffery, in an unpublished paper read at the AMS national meeting in 2003, suggested that the pedagogical use of ancient Greek accents may have contributed to the development of musical neumes.

<sup>7</sup> “The fact [is] that the Gregorian Chant tradition was, in its early centuries, an oral performance practice... The oral tradition was translated after the ninth century into writing. But the evolution from a performance practice represented in writing, to a tradition of composing, transmission, and reading, took place over a span of centuries.” (Treitler, 1982)

<sup>8</sup> “The church musicians who opted for the inexact aides-mémoire of staffless neumes – for skeletal notations that ignored exact pitch-heights and bypassed many nuances – were content with incomplete representations of musical substance because the full substance seemed safely logged in memory. This simple calculus of notation and memory says that the Gregorian chants from their first neumatism were no longer ‘improvised,’ that few if any options were left for the strategies and vagaries of individual performers. The chants were concretized reified entities, recognizable in their specific melodic dress, integrally stored and reproducible from memory.” (Levy, 1998).

## Melodic Contour Analysis Tool

Our tool takes in a (digitized) monophonic or heterophonic recording and produces a series of successively more refined and abstract representations of the melodic contours.

It first estimates the fundamental frequency (“F0,” in this case equivalent to pitch) and signal energy (related to loudness) as functions of time. We use the SWIPEP fundamental frequency estimator (Camacho 2007) with all default parameters except for upper and lower frequency bounds hand-tuned for each example. For signal energy we simply take the sum of squares of signal values in each non-overlapping 10-ms rectangular window.

The next step is to identify pauses between phrases, so as to eliminate the meaningless and wildly varying F0 estimates during these noisy regions. We define an energy threshold, generally 40 decibels below each recording’s maximum. If the signal energy stays below this threshold for at least 300 ms then the quiet region is treated as silence and its F0 estimates are ignored. Figure 1 displays an excerpt of the F0 and energy curves for Sura Al-Qadr Husari XXX. Provide more details XXX.

The next step is pitch quantization. Rather than externally imposing a particular set of pitches such as an equal-tempered chromatic or diatonic scale, we have developed a novel method for extracting a scale from an F0 envelope that is continuous (or at least very densely sampled) in both time and pitch. Our method is inspired by Krumhansl’s time-on-pitch histograms adding up the total amount of time spent on each pitch (Krumhansl 1990). We demand a pitch resolution of one cent<sup>9</sup>, so we cannot use a simple histogram.<sup>10</sup> Instead we use a statistical technique known as *nonparametric kernel density estimation*, with a Gaussian kernel.<sup>11</sup> The resulting curve is our density

estimate; like a histogram, it can be interpreted as the relative probability of each pitch appearing at any given point in time. Figure 2 shows this method’s density estimate given the F0 curve from Figure 1.

We interpret each peak in the density estimate as a note of the scale. We restrict the minimum interval between scale pitches (currently 80 cents by default) by choosing only the higher peak when there are two or more very close peaks. This method’s free parameter is the standard deviation of the Gaussian kernel, which provides an adjustable level of smoothness to our density estimate; we have obtained good results with a standard deviation of 30 cents. Note that this method has no knowledge of octaves.

Once we have determined the scale, pitch quantization is the trivial task of converting each F0 estimate to the nearest note of the scale.

In our opinion these derived scales are more true to the actual nature of pitch-contour relationships within oral/aural and semi-notated chant traditions. Instead of viewing these pitches to be deviations of pre-existing “normalized” scales our method defines a more differentiated scale from the outset. With our approach the scale tones do not require “normalization” and thereby exist in an autonomous microtonal environment defined solely on statistical occurrence of pitch within a temporal unfolding of the given melodic context.

## Interactive Web-Based Visualization and Exploration of Melodic Contours

We have developed a browsing interface that allows researchers to organize and analyze chant and the associated signs in a variety of ways. Each recording is manually segmented into units such as trope sign, neumes or words. The resulting segments can be rearranged in a variety of ways both manually and automatically. That way one can compare the beginning and ending pitches of any trope sign, neume or word or compare the relationships of one neume or trope sign to its neighbors. Our eventual goal is to explore the stability of melodic gesture and pitch content in a variety of contexts – within a given chant style and across chant styles. In addition we want to explore how the stability related to the chant texts and textual syntax. There are hard questions without clear answers. The user interface has been designed to assist and support the analysis conducted by expert musicologists without trying to impose a specific approach.

---

<sup>9</sup> One cent is 1/100 of a semitone, corresponding to a frequency difference of about 0.06%.

<sup>10</sup> F0 envelopes of singing generally vary by much more than one cent even within a steadily held note, even if there is “no vibrato.” Another way of thinking about the problem is that there isn’t enough data for so many histogram bins: if a 10-second phrase spans an octave (1200 cents) and our F0 envelope is sampled at 100 Hz then we have an average of less than one item per histogram bin.

<sup>11</sup> Thinking statistically, our scale is related to a probability distribution giving the relative probability of each possible pitch. We can think of each F0 estimate (i.e., each sampled value of the F0 envelope) as a sample drawn from this unknown distribution, so our problem becomes one of estimating the unknown distribution given the observations.

Being able to categorize melodic formulae in a variety of ways allows for a larger database of their gestural identities, their functionality to parse syntax as well their regional traits and relations. A better understanding of how pitch and contour helps to create gesture in chant might allow for a more comprehensive view of the role of gesture in improvised, semi-improvised and notated chant examples. The pitch contours of segments can be viewed at different levels of detail and smoothness using a histogram-based method.

We have chosen to implement the interface as a web-based Flash program. Web-based interfaces can increase the accessibility and usability of a program, make it easier to provide updates, and can enhance collaboration between colleagues by providing functionality that lets researchers more easily communicate their results to each other. The interface has four main sections: a sound player, a main window to display the pitch contours, a control window, and a histogram window.

The sound player window displays a spectrogram representation of the sound file with shuttle controls to let the user choose the current playback position in the sound file. It also provides controls to start and pause playback of the sound, to change the volume.

The main F0 display section, found above the sound player, is a window that shows all the pitch contours for the song as icons that can be repositioned automatically based on a variety of sorting criteria, or alternatively can be manually positioned by the user. At the top of each of these icons the name of the sign is shown and underneath the F0 contour is displayed. The shuttle control of the main sound player is linked to the shuttle controls in each of these icons, allowing the user to set the current playback state either by clicking on the sound player window, or directly in the icon of interest. When the user mouses over these icons, some salient data about the sign is displayed at the bottom of the screen.

The control window has a variety of buttons that control the sorting order of the icons in the main F0 display window. A user can sort the icons in playback order, alphabetical order, length order, and also by the beginning, ending, highest and lowest F0. The user can also display the sounds in an X-Y graph, with the x-axis representing highest F0 minus lowest F0, and the y-axis showing the ending F0 pitch minus the beginning F0 pitch. Also in this section are controls to toggle a mode to hear individual sounds when they are clicking

on, and controls to hide the pitch contour window leaving just the label. There are also buttons to choose if the original sound file, the sine wave representation, or the quantized sine wave representation is played back to the user.

When an icon in the main F0 display window is clicked, the histogram window shows a histogram of the distribution of quantized pitches in the selected sign. Below this histogram is a slider that can be dragged back and forth to choose how many of the top histogram bins will be used to generate a reduced dimensionality gesture representation of the F0 curve. In the limiting case of the maximum value, the gesture curve is exactly the quantized F0 curve. At lower values, only the histogram bins with the most items are used to draw the gesture curve, which has the effect of reducing the impact of outlier values and provide a more smooth “abstract” contour.

Multiple signs can be selected by shift-clicking. When more than one sign is selected, the histogram window includes the data from all the selected signs. This mode is generally used to select all variations of a particular sign, the gesture representation is then calculated using the sum of all the pitches found in that particular sign, enhancing the quality of the gesture representation.

Below the histogram window is a window that shows a zoomed in graph of the selected F0 contours. When more than one F0 contour is selected, the lines in the graph are color coded to make it possible to easily distinguish the different selected signs.

## Discussion and Future Work

The identity of chant formulae in oral/aural chant traditions is to a large extent determined by gesture/contour rather than by discrete pitches.

Computational approaches assist with the analysis of these gestures/contours and enables a the juxtaposition of multiple views at different levels of detail in a variety of analytical (paradigmatic and sytegmatic) contexts.

The possibilities for such complex juxtapositions of analysis methods would be difficult if not impossible without such computer assisted analysis.

Employing these tools we plan to examine melodic gesture in the interchange between oral and written cultures, between language as improvised speaking and language as spoken text, and between spoken and sung language.

While this analysis intends to shed light on how melodic gestures help to structure textual syntax, we will also continue to speculate about the historical development and of such functionality. Examples of chant as ritual process, chant as syntactical code and chant as “composed exegesis” have been used in order to explain the complex relationships between musical process, code and object.

## REFERENCES

Camacho, Arturo, 2007 *SWIPE: A Sawtooth Waveform Inspired Pitch Estimator for Speech and Music*. PhD Dissertation, University of Florida.

Karp, Theodore, 1998 *Aspects of Orality and Formularity in Gregorian Chant*. Evanston: Northwestern University Press.

Krumhansl, C. L. 1990 *Cognitive Foundations of Musical Pitch*. Oxford: Oxford University Press.

Levy, Kenneth. 1998. *Gregorian Chant and the Carolingians* Princeton: Princeton University Press 137.

Treitler, Leo. 1982 “The Early History of Music Writing in the West, *Journal of the American Musicological Society*, Volume 35, Chicago: University of Chicago Press, 237.

Wigoder Geoffrey, ed., et al. 1989, “Masora,” *The Encyclopedia of Judaism*. New York: MacMillan Publishing Company 468.