Decoding the Song: Histogram-Based Paradigmatic and Syntagmatic Analysis of Melodic Formulae in Hungarian Laments, Torah Trope, Tenth Century Plainchant and Koran Recitation

Dániel Péter Biró[1], Steven Ness[2],W. Andrew Schloss[1], George Tzanetakis[2], Matthew Wright[1,2]
University of Victoria, School of Music[1] and Department of Computer Science[2]
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: Hungarian *siratók* (laments), Torah cantillation, tenth 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's melodic framework is governed by the particular religious context for performance.

The *sirató* is a lament ritual from Hungary that goes back at least to the Middle Ages.[1] This improvised song type is integral for our study, as it exemplifies inherent relationships between speech and singing while demonstrating stable melodic formulae within an oral/aural ritual context.

Jewish Torah trope is "read" using the twenty-two cantillation signs of the *te'amei hamikra,*[2] developed by the Masorete rabbis[3] between the sixth to the ninth centuries. The melodic formulae of Torah trope govern syntax, pronuciation and meaning. While the written *te'amim* have not changed since the tenth century C.E., their corresponding melodic formulae are determined not only by Jewish tradition of cantillation but also by the melodic framework of their surrounding musical environment.

The performance framework for Koran recitation is not determined by text or by notation but by rules of recitation that are primarily handed down orally (Zimmermann 2000, p. 128).[4] Here the hierarchy of spoken syntax, expression and pronunciation play a major role in determining the vocal styles of *Tajwīd*[5] and *Tartīl*[6]. The resulting melodic phrases, performed not as "song" but "recitation" are, like those of Torah trope, determined by both the religious and larger musical cultural contexts.

---

[1] Lamenting by women was common already in Biblical times: "Mourning songs for the dead also go back to primitive times. Although every religion and secular form of legislation… has endeavored to control mourning practices, they are still customary even today" (Kodály 1960, p. 76).

[2] The term "ta'amei hamikra" means literally "the meaning of the reading."

[3] "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" (Wigoder 1989, p. 468).

[4] "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' ãn) could be 'the recited'" (Zimmerman 2000, p. 27, translation by Biró).

[5] "*Tajwīd* [is] the system of rules regulating the correct oral rendition of the Qur'an. The importance of *Tajwīd* to any study of the Qur'an cannot be overestimated: *Tajwīd,* preserves the nature of a revelation whose meaning is expressed as much as by its sound as by it content and expression, and guards it from distortion by a comprehensive set of regulations which govern many of the parameters of the sound production, such as duration of syllable, vocal timbre and pronunciation" (Nelson 1985, p. 14). <http://www.grovemusic.com.ezproxy.library.uvic.ca>

[6] "*Tartīl,* another term for recitation, especially implies slow deliberate attention to meaning, for contemplation." (Neubaurer and Doublday).

The early plainchant neumes came from a logogenic culture that was based on textual memorization; the singing of memorized chants was central to the preservation of a tradition that developed over centuries (Treitler 1982). 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. [7]

Chant scholars have investigated historical and phenomenological aspects of chant formulae to discover how improvised melodies might have developed to become stable melodic entities, paving the way for the development of notation.[8] A main aspect of such investigations has been to 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.

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 100 ms then the quiet region is treated as silence and its F0 estimates are ignored. Figure 1 shows an excerpt of the F0 and energy curves for an excerpt from the Koran *sura* ("section") *Al-Qadr* ("destiny") recited by the renowned Sheikh Mahmûd Khalîl al-Husarî from Egypt.

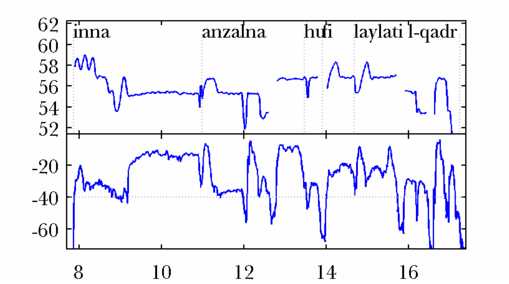

Figure 1: Pitch (top, MIDI units) and Energy (bottom, decibels) Contours

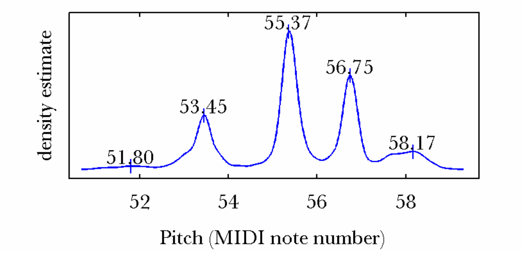

Figure 2: Recording-specific scale derivation

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 (Krumhanl 1990). We demand a pitch resolution of one cent[9], so we cannot use a simple histogram.[10] Instead we use a statistical technique known as *nonparametric*

[7] "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, p. 237).

[8] "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" (Levy 1998, p. 137).

[9] One cent is 1/100 of a semitone, corresponding to a frequency difference of about 0.06%.

[10] 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.

kernel density estimation, with a Gaussian kernel.[11] 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 musical 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 segments in a variety of ways. The user manually segments each recording into the appropriate units for each chant type (such as trope sign, neumes, semantic units, or words). The pitch contours of these segments can be viewed at different levels of detail and smoothness using a histogram-based method. The segments can also 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.

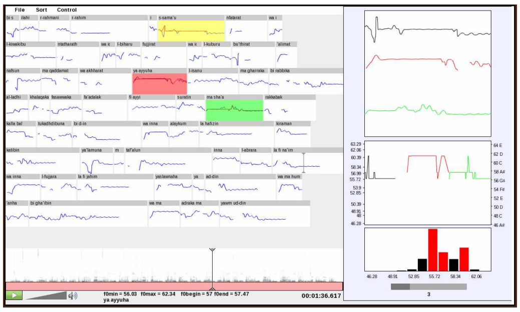

Figure 3: Screen-shot of interface

Our eventual goal is to explore the stability of melodic gesture and pitch content in a variety of contexts both 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 design to assist and support the analysis conducted by expert musicologists without trying to impose a specific approach. 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, and 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.

We have chosen to implement the interface as a web-based Flash program, which can be accessed at <http://cantillation.sness.net>. Web 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 (shown in Figure 3) 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 window* 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

[11] Thinking statistically, our scale is related to a 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.

positioned by the user. The name of each segment (from the initial segmentation step) appears above its F0 contour. 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 allowing the user to choose to hear the original sound file, the F0 curve applied to a sine wave, or the quantized F0 curve applied to a sine wave.[12]

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 to choose how many of the largest histogram bins will be used to generate a simplified contour representation of the F0 curve. In the limiting case of selecting *all* histogram bins, the reduced curve is exactly the quantized F0 curve. At lower values, only the histogram bins with the most items are used to draw the reduced curve, which has the effect of reducing the impact of outlier values and providing a smoother "abstract" contour.

Shift-clicking selects multiple signs; in this case the histogram window includes the data from all the selected signs. We often select all segments with the same word, trope sign, or neume; this causes the simplified contour representation to be calculated using the sum of all the pitches found in that particular sign, enhancing the quality of the simplified contour 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 the juxtaposition of multiple views at different levels of detail in a variety of analytical (paradigmatic and syntagmatic) contexts.

The possibilities for such complex analysis methods would be difficult if not impossible without such computer-assisted analysis. Employing these tools we hope to better understand the role of and interchange between melodic formulae in oral/aural and written chant cultures. While our present analysis investigates melodic formulae primarily in terms of their gestural content and semantic functionality, we hope that these methods might allow scholars to reach a better understanding of the historical development of melodic formulae within various chant traditions.

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.

Kodály, Zoltán, 1960. *Folk Music of Hungary*. Budapest: Corvina Press.

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

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

Nelson, Kristina, 1985. *The Art of Reciting the Koran,* Austin, University of Texas Press.

Neubaurer, Eckhard and Veronica Doubleday: 'Qur'anic recitation,' *Grove Music Online* ed. L. Macy (Accessed 6 June 2008), <http://www.grovemusic.com>

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.

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

Zimmermann, Heidi *Tora und Shira: Untersuchungen zur Musikauffassung des rabbinischen Judentums.* Bern: Peter Lang, 2000, 128.

[12] The sine waves also follow the computed energy curves.