

A COMPUTATIONAL INVESTIGATION OF MELODIC CONTOUR STABILITY IN JEWISH TORAH TROPE PERFORMANCE TRADITIONS

Peter van Kranenburg
Meertens Institute

peter.van.kranenburg@meertens.knaw.nl

Dániel Péter Biró
University of Victoria

dpbiro@finearts.uvic.ca

Steven Ness, George Tzanetakis
University of Victoria

sness@sness.net, gtzan@cs.uvic.ca

ABSTRACT

The cantillation signs of the Jewish Torah trope are of particular interest to chant scholars interested in the gradual transformation of oral music performance into notation. Each sign, placed above or below the text, acts as a “melodic idea” which either connects or divides words in order to clarify the syntax, punctuation and, in some cases, meaning of the text. Unlike standard music notation, the interpretations of each sign are flexible and influenced by regional traditions, practices of given Jewish communities, larger musical influences beyond Jewish communities, and improvisatory elements incorporated by a given reader. In this paper we describe our collaborative work in developing and using computational tools to assess the stability of melodic formulas of cantillation signs based on two different performance traditions. We also show that a musically motivated alignment algorithm obtains better results than the more commonly used dynamic time warping method for calculating similarity between pitch contours. Using a participatory design process our team, which includes a domain expert, has developed an interactive web-based interface that enables researchers to explore aurally and visually chant recordings and explore the relations between signs, gestures and musical representations.

1. INTRODUCTION

In the last ten years there has been a growing interest in music information retrieval (MIR). A variety of techniques for automatically analyzing music based on both symbolic and audio representations have been developed. In most cases the target user of MIR systems has been the average music listener rather than the specialist. There is an even longer tradition of computational musicology dating back to the 1950s of using mathematics, statistics and eventually

computers to study music. Most of this work in computational musicology has focused on the symbolic domain and western music notation. More recently the idea of Computational Ethnomusicology in which MIR techniques are used to support research in musics from around the world has been proposed [9]. Audio analysis techniques can be used for empirical research on field recordings for which no transcription is available or feasible.

The study of religious chant is of particular interest to musicologists as it can help understand the transition from oral transmission to codified notation. Jewish Torah trope is “read” using the thirty cantillation signs of the *te’amei hamikra*, developed by the Masoretic School between the sixth to the tenth centuries. The Masoretes concurrently inscribed the *te’amim* along with the vowels of the Hebrew letters in order to ensure accuracy in future Torah reading, thereby altering the previous mode of oral transmission. 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 *te’amim* consist of thirty graphic signs. Each sign, placed above or below the text, acts as a “melodic idea,” which either melodically connects or divides words in order to make the text understandable by clarifying syntax, pronunciation and, in some cases, musical meaning. The signs serve to indicate the melodic contour of a given melody. Although the thirty signs of the *te’amim* are employed in a consistent manner throughout the Hebrew Bible, their interpretation is flexible: each sign’s modal structure and melodic gesture is determined by the text portion, the liturgy, by regional traditions as well as by improvisatory elements incorporated by a given “reader”.

In the liturgical performance, the *ba’al koreh* (‘the owner of reading’) embellishes the text with a melodic code, providing the framework to decode the textual syntax of the read Torah text by the reading religious community, for whom text, and not melody, is primary. Since their inscription, the primary functionality of the *te’amim*, to

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page.

© 2011 International Society for Music Information Retrieval.

structure pronunciation and syntax, remained intact. But as the Jewish people were dispersed throughout the world, secondary levels of musical code were incorporated into the *te'amim*. Borrowed melodies and modal structures, taken from surrounding musical cultures allowed not only for new melodic interpretations but also for external *semiotic musical meaning* to permeate the musical interpretation of the text. As an example, the left part of Figure 1 shows the sign for the *etnachta* and the right part shows the melodic contour of the performance of an *etnachta*.

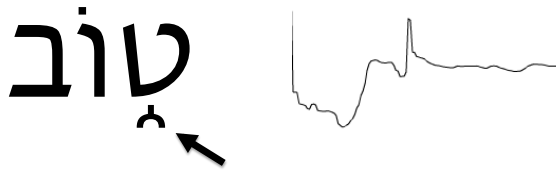


Figure 1. The notational sign of the *etnachta* (indicated by the arrow), and the melodic contour of an *etnachta*.

Chant scholars have investigated historical and phenomenological aspects of melodic formulas within Jewish Torah trope in order to discover how improvised melodies might have developed to become stable melodic entities in given Jewish communities. In this paper we investigate how computational approaches can be used to support research in this area. More specific, audio analysis is combined with content-based similarity retrieval to explore the ways in which melodic contour defines melodic identities. In particular the question of melodic stability is investigated. Observing certain key *te'amim* such as *etnachta* and *tifha* we investigate aspects of self-similarity within Torah trope within and across various Jewish communities (based on recordings of Hungarian and Moroccan Torah trope). This might give us a better sense of the role of melodic gesture in melodic formulae in Jewish Torah trope practice and possibly a new understanding of the relationship between improvisation and notation-based chant in and amongst these divergent traditions.

It is also possible that some of the *te'amim* have precursors to music (for instance basic syntactical divisions, exclamations and sentence cadence structures). The actual performance of the *te'amim* also points to musical aspects that, as scholars have pointed out, were coming from musical cultures outside of Judaism (see e.g., [2]). That which has been historically studied, the relationship between Ashkenazi Torah trope and Christian plainchant, can now be tested in terms of musical data analysis. By measuring the flexibility and variability of the *te'amim* we can show how fixed musical structures and improvisation within these traditions co-exist.

2. RELATED WORK

Although most of existing work in music information retrieval has focused on either classical music or modern popular music, in recent years there has been a growing interest in applying MIR techniques to other music traditions. The term Computational Ethnomusicology [9] has been used to describe such work. There are both challenges and opportunities in applying MIR techniques to ethnic music [5]. Some representative examples include: classification of raag using pitch class distributions [4], comparative analysis of western and non-western traditions using automatic tonal feature extraction [8], rhythmic similarity applied to greek and african traditional music [10], and singer identification in rembetiko music [1].

The goal of this project is develop tools to study Torah cantillation [14]. Of particular interest is the influence of outside music cultures such as christian plainchant to the performance of Jewish Torah trope [2]. The primary method that has been used in the past to study chant recordings has been listening and manual annotation. We believe that the combination of automatic analysis with web-based interactive visualizations can open new possibilities in empirical musicological analysis of chant recordings. In the development of both our techniques and web-based interface we have followed an iterative participatory design process where the domain expert (one of the authors) has been regularly providing feedback and suggestions. Our approach is based on ideas from the field of query-by-humming (QBH) [6, 7] adapted to the particular characteristics and constraints of our domain. In previous work [12] we compared various representations and methods of quantizing pitch contours in various chant traditions using a similarity retrieval paradigm. In this paper we focus on Jewish Torah trope, propose an alternative alignment method and show how the developed techniques can be used to inform musicological inquiries. To the best of our knowledge a data-rich approach to the study of Torah trope as presented in this paper has not been attempted before.

3. DATA ANALYSIS

For this small-scale study, we use the recordings of two readings of the same Torah passage, one from the Hungarian (Ashkenazi) tradition¹ and the one from the Moroccan (Sephardic) tradition. The two recordings used in this study can be consulted at: <http://cantillion.sness.net/ismir2011>. Both recordings have manually been segmented into the individual *te'amim* by the author who is a domain expert. Even though we considered the possibility

¹ Recordings used with permission of the Feher Music Center in Tel Aviv, Israel. Although this version was catalogued as being an example of Hungarian cantillation, the trope melody and pronunciation correspond more to Italian practices of Torah trope.

of creating an automatic segmentation tool, it was decided that the task was too subjective and critical to automate. Each segment is annotated with a word/symbol that is related to the corresponding cantillation sign. Each recording contains approximately 130 realizations of each *ta'am* with a total of 12 unique *te'amim*.

3.1 Pitch Contour Representation

Each recording has been converted to a sequence of frequency values using the SWIPEP fundamental frequency estimator [3] by estimating the fundamental frequency in non-overlapping time-windows of 10ms. The frequency sequences have been converted to sequences of real-valued MIDI pitches with a precision of 1 cent (which is 1/100 of an equally tempered semitone, corresponding to a frequency difference of about 0.06%). By allowing real-valued pitches we have a one-to-one correspondence to the frequencies, and a linear scale in the pitch domain. For each of the recordings, we derive a melodic scale by detecting the peaks in a non-parametric density estimation of the distribution of pitches, using a Gaussian kernel. This can be viewed as a smoothed frequency histogram. Prominent peaks in the histogram correspond to salient pitches and can be used to form a discrete pitch scale that is specific to the recording rather than any particular tuning system.

In a previous study [12], mean average precision values were computed for each of the scales containing 1 to 13 pitches, taking all realizations of the same *ta'am* as the query segment as relevant items, and using a distance measure based on dynamic time warping. The finding was that quantizing the melodic contours according to the scale containing *two* pitches resulted in the highest mean average precision. Apparently, the two most prevalent pitches have structural meaning.

In the current study we use a different approach. Instead of quantizing the melodic contours, we scale them linearly according to the two most prevalent pitches in the entire recording. We denote the higher and lower of the two prevalent pitches as p_{high} and p_{low} , respectively. Each pitch is scaled relative to p_{low} in units of the difference between p_{high} and p_{low} . Thus, scaled pitches with value < 0 are below the lowest of the two prevalent pitches and pitches with value > 1 are above the highest of the two and pitches between 0 and 1 are between the two prevalent pitches. As a result, different trope performances, sung at different absolute pitch heights, are comparable.

3.2 A distance measure for melodic segments

On the acquired scaled pitch contours we apply an alignment algorithm as described in [13], interpreting the alignment score as similarity measure. This approach is closely related to the use of dynamic time warping in [12], but

the current approach uses a more advanced, musicologically informed, scoring function for the individual elements of the pitch sequences.

We use the Needleman-Wunsch global alignment algorithm [11]. This algorithm finds an optimal alignment of two sequences of symbols, which, in our case, are sequences of pitches. The quality of an alignment is measured by the alignment score, which is the sum of the alignment scores of the individual symbols. If we consider two sequences of symbols $\mathbf{x} : x_1, \dots, x_i, \dots, x_n$, and $\mathbf{y} : y_1, \dots, y_j, \dots, y_m$, then symbol x_i can either be aligned with a symbol from sequence \mathbf{y} or with a gap. Both operations have a score, respectively the substitution score and the gap score. The gap score is mostly expressed as penalty, i.e. a negative score. The optimal alignment and its score are found by filling a matrix D recursively according to:

$$D(i, j) = \max \begin{cases} D(i-1, j-1) + S(x_i, y_j) \\ D(i-1, j) - \gamma \\ D(i, j-1) - \gamma \end{cases}, \quad (1)$$

in which $S(x_i, y_j)$ is a similarity measure for symbols, γ is the gap penalty, $D(0, 0) = 0$, $D(i, 0) = -i\gamma$, and $D(0, j) = -j\gamma$. $D(i, j)$ contains the score of the optimal alignment up to x_i and y_j and therefore, $D(m, n)$ contains the score of the optimal alignment of the complete sequences. We can obtain the alignment itself by tracing back from $D(m, n)$ to $D(0, 0)$; the standard dynamic programming algorithm has both time and space complexity $O(nm)$.

The similarity measure for symbols, which returns values in the interval $[-1, 1]$, is in our case defined as:

$$S(x, y) = \begin{cases} 1 - 4|sp_x - sp_y| & \text{if } |sp_x - sp_y| \leq 0.5 \\ -1 & \text{otherwise} \end{cases},$$

in which scaled pitch of symbol x is

$$sp_x = \frac{p_x - p_{low,x}}{p_{high,x} - p_{low,x}},$$

in which p_x is the pitch of symbol x , represented in continuous midi encoding, and $p_{low,x}$ and $p_{high,x}$ are the lowest and highest pitch in the entire recording to which symbol x belongs. sp_y is computed in the same way. We use a linear gap penalty function with $\gamma = 0.6$.

Since the score of an alignment depends on the length of the sequences, normalization is needed to compare different alignment scores. The alignment of two long sequences results in a much higher score than the alignment of two short sequences. Therefore, we divide the alignment score by the length of the shortest sequence. Thus, an exact match results in score 1, which is the maximal score. The scores are converted into distances by taking one minus the normalized score, resulting in distances greater than or equal to zero.

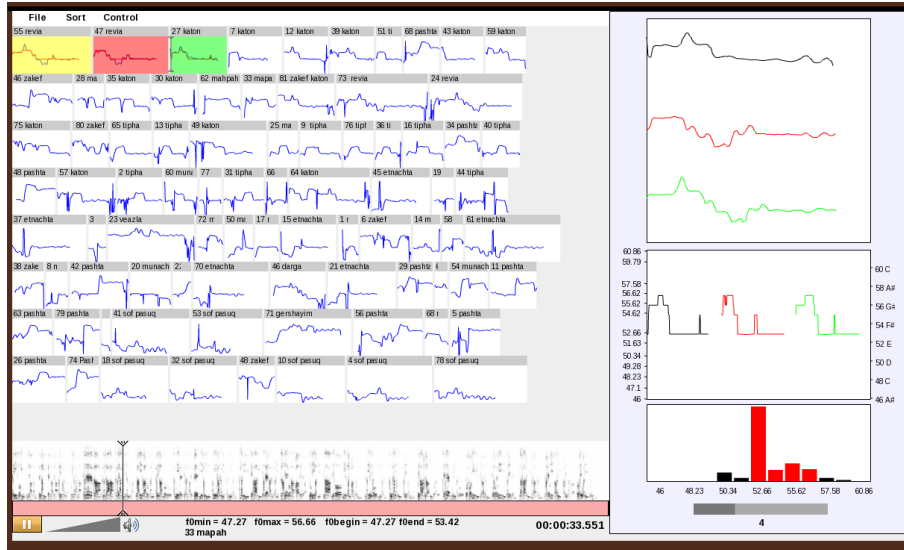


Figure 2. Web-based visualization interface which allows users to listen to audio, see pitch contour visualization of different signs, and to enable interactive similarity-based querying.

4. USER INTERFACE

We have developed a browsing interface that allows researchers to organize and analyze chant segments in a variety of ways. Each recording is manually segmented into *te'amim*. The pitch contours of these segments can be viewed at different levels of detail. They can also be rearranged in a variety of ways both manually and automatically.

The interface shown in Figure 2 has four main sections: a sound player, a main window to display the audio segments, 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. The main window shows all the segments of the recording as icons that can be repositioned automatically based on a variety of sorting criteria, or alternatively can be manually 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 way.

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 segment. 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 outliers and providing a smoother “abstract” contour. Shift-clicking selects multiple segments; in this case the histogram window includes the data from all the selected segments. We often select all segments with the same word or *ta'am*; this causes the simplified contour representation to be calculated using the sum of all the pitches found in that particular *ta'am*, enhancing the quality of the simplified contour representation. Figure 2 shows a screenshot of the browsing interface. We have implemented a mode that allows the researcher to sort the segments based on the alignment score from one segment to the other. The interface allows the user to select an arbitrary segment from the interface, and then perform a sorting of all other segments to it. In the example shown in Figure 2, the user has chosen a *revia*, and has sorted all the other segments based on their alignment-based distance from this first *revia*. One can see that the segment closest to this *revia* is another *revia* from a different section of the audio file.

5. RESULTS AND INTERPRETATION

To investigate the stability in performance of the various *te'amim*, we use two approaches. Firstly, we compute the mean average precision for each of the *te'amim* based on the alignment-distance. Each segment is taken as query and all renditions of the same *ta'am* are taken as relevant items. The higher the mean average precision, the higher the relevant items are on the ranked lists that are obtained by sorting all segments according to the distance to the query segment.

Ta'am (Morocco)	Average Precision (Morocco)	Ta'am (Hungary)	Average Precision (Hungary)
sofpasuq	0.550	sofpasuq	0.994
katon	0.399	revia	0.967
tipha	0.306	etnachta	0.945
mapah	0.299	pashta	0.683
pashta	0.269	tipha	0.673
revia	0.245	katon	0.562
etnachta	0.234	mapah	0.550
zakef	0.206	merha	0.530
merha	0.158	zakef	0.231
munach	0.147	munach	0.179
kadma	0.036	kadma	0.040

Table 1. Mean average precision for different te'amim based on the alignment distances.

The values are shown in Table 1.

Secondly, we show the distribution of distances between renditions of the same *ta'am* by plotting histograms of those distances. Figure 3 shows the distribution of alignment-based distances between unrelated segments. This histogram can be used as reference for comparing distances between related segments. The interface, as described in the previous section, is used to examine the relations between individual audio segments.

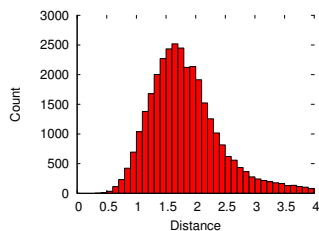


Figure 3. Distribution of distances between unrelated segments.

The obtained overall mean average precisions are 0.644 for the Hungarian rendition and 0.309 for the Moroccan one, which are improvements concerning the results that were previously achieved in [12] (0.505 and 0.229 respectively). Using the current alignment-approach, the segments are better recognized, but the overall trend appears the same, namely a better retrieval result for the Hungarian rendition as compared to the Moroccan. Since we do not know a priori whether every *ta'am* has a high level of distinction, we cannot draw conclusions about the quality of our distance measure from the MAP-values. A low MAP-value does not necessarily mean that the distance measure fails, but could also indicate that the performance of the specific *ta'am* is

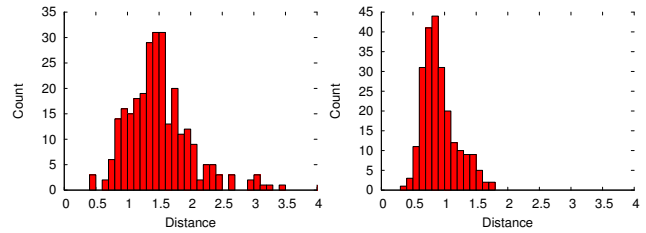


Figure 4. Distribution of distances between renditions of the *tipha* in the Moroccan interpretation (left) and the Hungarian interpretation (right).

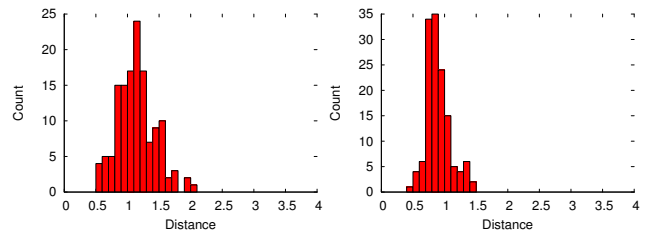


Figure 5. Distribution of distances between renditions of the *sofpasuq* in the Moroccan interpretation (left) and the Hungarian interpretation (right).

variable or not distinct from performance of other *te'amim*.

Therefore, in remainder of our analysis, we focus on various key *te'amim*, using differences between distances and mean-average-precisions, along with musicological domain knowledge, to draw conclusions. Observing the renditions of *sofpasuq* and *tipha* in the Hungarian tradition, one can derive that they inhibit a definite melodic stability. For the *sofpasuq* we obtain a mean average precision as high as 0.994 and for the *tipha* 0.673 (for comparison, the figures for the Moroccan performance are 0.550 and 0.306 respectively). This indicates that the 17 *sofpasus* are both similar to each other and distinct from all other *te'amim*. The same applies to a somewhat lesser extent to the 24 *tiph*as. These findings are confirmed by the distributions of distances as shown in Figures 4 and 5.

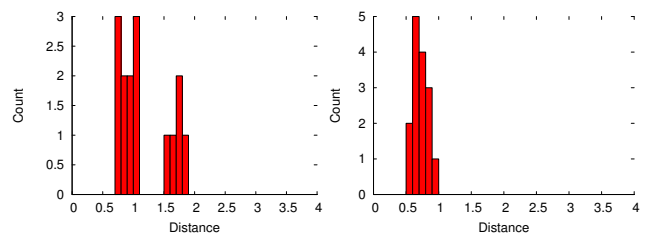


Figure 6. Distribution of distances between renditions of the *etnachta* in the Moroccan interpretation (left) and the Hungarian interpretation (right).

Analyzing the distribution of distances between Moroc-

can renditions of the *etnachta*, as shown in the left histogram in Figure 6, one finds increased melodic variation while the Hungarian interpretation shows greater melodic stability. This is significant, as *etnachta* is an example of a disjunctive *ta'am*, that has a clear functionality as a syntactical divider within a given sentence. Such a melodic stability might have been due to the influence of Christian Chant on Jewish communities in Europe, as is the thesis of Avenary [2]. Simultaneously, our approach using two structurally important pitches also corresponds to the possible influence of recitation and final tone as being primary tonal indicators within Askenazi chant practice (which the Hungarian Torah Trope is part of), thereby allowing for a greater melodic stability per trope sign than in Sephardic chant.

The findings are interesting when observed in connection with musicological and music historical studies of Torah trope. It has long been known that the variety of melodic formulae in Ashkenazi trope exceeded that of Sephardic trope renditions. The *te'amim* actually entail more symbols than necessary for syntactical divisions. That being said, in certain *te'amim*, like in the version of *etnachta*, a greater amount of melodic variability is presented. This is not mirrored in the example of *tipha*, which serves to combine words to make a clear syntactical unit. In both Hungarian and Moroccan variants this *ta'am* shows a greater degree of stability. This shows that certain conjunctive *te'amim*, which show greater melodic stability, might also act as more stable syntactical anchors in both traditions. One might investigate if this is also true in other traditions (Iranian, Yemenite and Lithuanian).

6. FUTURE WORK

In the current study, we took the two most prevalent pitches for scaling. There are reasons to assume that for various performance traditions different numbers of pitches are of structural importance. We will investigate this in future research. The presented method proves useful for the two recordings under investigation. In a next stage, we will collect much more data, with the aim to study stability and variation between and within performance traditions of Torah trope on a large scale, integrating the results into ongoing musicological and historical research on this topic.

7. REFERENCES

- [1] Y. Stylianou A. Holzapfel. Signer identification in rembetiko music. In *Proc. Sound and Music Computing (SMC)*, 2009.
- [2] H. Avenary. *The Ashkenazi Tradition of Biblical Chant Between 1500 and 1900. Tel-Aviv and Jerusalem*. Tel-Aviv University, Faculty of Fine Arts, School of Jewish Studies, 1978.
- [3] A. Camacho. *A Sawtooth Waveform Inspired Pitch Estimator for Speech and Music*. PhD thesis, University of Florida, 2007.
- [4] P. Chordia and A. Rae. Raag recognition using pitch-class and pitch-class dyad distributions. In *Proc. Int. Conf. on Music Information Retrieval (ISMIR)*, 2007.
- [5] O. Cornelis, M. Lesaffre, D. Moelants, and M. Leman. Access to ethnic music: Advances and perspectives in content-based music informatino retrieval. *Signal Processing*, 90(4):1008–1031, 2010.
- [6] R. Dannenberg, W.P. Birmingham, B. Pardo, N. Hu, C. Meek, and G. Tzanetakis. A comparative evaluation of search techniques for query-by-humming using the musart testbed. *J. Am. Soc. Inf. Sci. Technol.*, 58(5):687–701, 2007.
- [7] A. Ghias, J. Logan, D. Chamberlin, and B.C. Smith. Query by humming: musical information retrieval in an audio database. In *Proc. ACM Int. Conf. on Multimedia*, pages 231–236, 1995.
- [8] E. Gomez and P. Herrera. Comparative Analysis of Music Recordings from Western and non-western traditions by Automatic Tonal Feature Extraction. *Empirical Musicology Review*, 3(3):140–156, 2008.
- [9] G.Tzanetakis, A. Kapur, W.A. Schloss, and M. Wright. Computational ethnomusicology. *Journal of interdisciplinary music studies*, 1(2):1–24, 2007.
- [10] A. Pikrakis I. Antonopoulos, S. Theodoridis, O. Cornelis, D. Moelants, and M. Leman. Music retrieval by rhythmic similarity applied on greek and african traditional music. In *Proc. Int. Conf. on Music Information Retrieval (ISMIR)*, 2007.
- [11] Saul B. Needleman and Christian D. Wunsch. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, 48(3):443–453, 1970.
- [12] Steven R. Ness, Dániel Péter Biró, and George Tzanetakis. Computer-assisted cantillation and chant research using content-aware web visualization tools. *Multimedia Tools Appl.*, 48(1):207–224, 2010.
- [13] P. Van Kranenburg, A. Volk, F. Wiering, and R.C. Veltkamp. Musical models for folk-song melody alignment. In *Proc. Int. Conf. on Music Information Retrieval (ISMIR)*, pages 507–512, 2009.
- [14] Heidi Zimmermann. *Untersuchungen zur Musikauffassung des rabbinischen Judentums*. Peter Lang, Bern, 2000.