AUTOMATIC ANALYSIS OF MUSIC: PERFORMANCE OF CANTILLATION SIGNS IN YEMENITE JEWISH TRADITIONAL CANTILLATION

Adiel Ben-Shalom[1], Joseph Keshet[2], Roni Yeger-Granot[1]

[1] *Musicology Department, Hebrew University*

[2] *Department of Computer Science, Bar-Ilan University*

Correspondence should be addressed to: adiel.benshalom@gmail.com

Abstract: **Jewish cantillation is a ritual chanting of readings from the Hebrew bible in the synagogue services. The chants are written using special signs or marks printed in the Hebrew Bible. The purpose of the cantillation signs is to guide the chanting of the sacred texts during public worship, and to clarify the syntactical structure of the text while the specifics of the performance serve in addition as a rehtorical device and as a commentary to the text itself, highlighting important or affective points in the text. Musical analysis of the cantillation requires a detailed estimation of the distributions of each of the cantillation signs. This can be achieved by manual annotation of the audio, but such a method is subjective and labor-intensive. Using high-precision computational methods we developed a framework for automatic extraction of cantilation signs performance from recordings and we show an objective, automatic analysis of one of the cantillation signs, *Sof-Pasuq* (end of verse). We show that using variations on the finalis pitch, the reader divides the text into sections of coherent meaning, and by doing so he provides some subjective interpretation to the text. This finding is quite interesting because it demands an exceptional memory for pitch on the part of the reader, as well as presumably by the audience. We analyzed automatically two different readers of the bible, and give quantitative comparison between their chantings.**

1. INTRODUCTION

In the Jewish tradition there is a custom to read a portion of the Torah (also referred often as pentateuch), the first five books of the Hebrew bible, during the liturgical event in the synagogue every Saturday. The reading is a ritual singing of the text in a traditional way that is marked using special signs printed below, or above the letters. Usually one sign for each word. An example of cantillation signs is presented in Figure 1.

The use of the cantillation signs dates back to at least the 9th-10th century. The performance practice of each sign was passed on orally along the years, with the actual musical realization of each sign different among different Jewish communities, depending on their origin. Thus, for example the musical realization of the cantillation sign of Jews originated from Germany is different from the chant of Jews from Morocco or Tunisia.

Besides their use as an introduction to the reader how to read (or chant) each word, the cantillation signs also provide a detailed syntactic parsing of the text [1, 2], as well as an indication as to the exact stress location in each word. A comprehensive musical analysis of the cantillation was done in a pioneering work by ethnomusicologist Abraham. Z. Idelsohn who conducted a comparative study on the different musical realizations of the cantillation signs among Jewish communities [3]. Later studies by ethnomusicologists focused on musical analysis of specific traditions (e.g. [4, 5])

However there is almost no research on computational and acoustic analysis of Jewish cantillation. Such an analysis requires a detailed estimation of the distributions of each of the cantillation signs. This can be achieved by manual annotation of the audio, but such a method is subjective, and labor-intensive. Recently, a research work on Torah cantillation signs was done done by Kranenburg et al., [6]. In this work, the authors proposed a framework for computational investigation of the cantillation signs, checking the stability of melodic formulas of cantillation signs based on two different performance traditions. The audio data was manually segmented by a domain expert and the cantillation signs acoustical information was extracted using this segmentation. A recent work on computation analysis of classical Turkish music proposes to use phoneme alignement algorithm, trained on Turkish singing voice, as a framework for further analysis [7]

Figure 1: Genesis 1:9: God said, "Let the waters be collected." Letters in black, vowel points and d'geshim (gemination marks) in red, cantillation in blue (source: Wikipedia).

In the current work we propose to use high-precision computational methods for objective, automatic analysis of one of the cantillation signs. The method is based on machine learning algorithms originally designed for speech analysis, and adapted in this work to handle the biblical text, and to handle singing rather than read speech. In this preliminary work, we analyzed only one of the cantillation signs, *Sof Pasuq* (end of verse), and compared two distinguished readers of the Yemenite tradition, who read the same text. We present a quantitative comparison between their chantings, that was performed automatically.

The paper is organized as follows. In the next section we describe the corpus of the audio we used for training and for analysis. In Section 3 we present the automatic tools we used, and the way they were adapted to handle the biblical text and the cantillation chant. In Section 4 we outline quantitative results and conclude in Section 5.

2. CORPUS

Our corpus contains two distinguished readers of the Hebrew bible, reading in the Yemenite tradition. The recordings were taken from *The Jewish Oral Traditions Research Center* in the Hebrew University. These recordings were taken between 1955-1965, a period, in which a huge effort was done by linguistics researchers to record all the different dialects of Hebrew from Jews who came from the diaspora. All the recordings were originally stored on tapes and digitized few years ago.

The readers that we chose, Rabbi Shlomo Kare and Rabbi Shalom Cohen both born in Sana'a, Yemen and considered to be reliable informants of Jewish Yemenite tradition from Center of Yemen. Both of them recorded the entire Torah reading, although not all the material was digitized yet.

3. AUTOMATIC ANALYSIS

In this section we describe the algorithmic steps of the automatic analysis of the text and audio. First, we converted the biblical text into phonemes. This was done by automatic mapping of rules that takes into account the vowel points written in Hebrew. Then we aligned the phonemes to the audio by a modified version of an accurate state-of-the-art phoneme aligner, which was adapted to chant audio. We extracted the audio portions that correspond to the cantillation sign *Sof Pasuq* and analyze its pitch contour and distribution.

3.1. Mapping the Biblical Text into Phonemes

Hebrew is written without vowels. The word *yeled* (child) is written with the Hebrew equivalent of *y*, *l* and *d*. It is impossible to know how to pronounce a word from the way it is written, and actually the word witten as *y*, *l* and *d* can also be read as *yalad* (gave birth) or as *yiled* (to deliver a baby). Around the 9th century, a system for indicating vowels by using small dashes and dots placed below or above the consonant letters was developed (red signs in Figure 1), and it is used to write the biblical text, as well as in modern Hebrew.

We developed a procedure for the automatic conversion of dotted Hebrew text into its phonetic content. This conversion procedure is based on a set of rules that were manually compiled. Consonants and vowels are mapped to their phonemic realization either instantly or by looking ahead or back to the previous letter or vocalization. Special rules based on the sonority level of consonants and vowels were created to predict whether the vocalization Sheva (Shva) at the beginning of a word is realized as quiescent Sheva (Shva Nah) or mobile Sheva (Shva Na'). Results shows that the phonemic transcriptions from this automatic procedure are on a par with the manual phonemic transcription of the words in the corpus. This tool allows us to use automatically aligned Hebrew dotted orthographic text with the speech.

Our algorithm takes a reliable digital version of the biblical text that contains the cantilation signs [8]. We extract along the phoneme information the exact location of the cantilation signs in the text.

3.2. Phoneme Classifier

A basic tool in speech and language processing is a phoneme classifier. Given a fixed frame of speech (10 msec), the phoneme classifier returns the most probable phoneme that was produced in that frame. More specifically, it returns a vector of phonemes scores. The highest score in this vector is the prediction of the most probable score.

We used multiclass Passive-Aggressive algorithm [9], which is aimed at minimizing the misclassification error rate. We trained it on an manually annotated aligned data which contains Chapter 5 from Genesis chanted by Kare. We randomly chose 60% of the verses in the chapter to serve as the training data, 20% as a development set and use the rest as a test set. The manual phoneme annotation was done by linguistic expert. We extracted standard 12 MFCC features and log energy with their deltas and double deltas to form 39-dimensional acoustic feature vectors. The window size and the frame size were 25 msec and 10 msec, respectively. We normalized the feature to have zero mean and standard deviation of 1. Then, we applied RBF kernel approximation as in [10] with a parameter that was chosen on a development set to be $\sigma^2 = 19$. We assessed the performance of the classifier on the test set, and got error rate of 26.8%, which is considered to be relatively high.

3.3. Phoneme Aligner

Phoneme alignment is the task of proper positioning of a sequence of phonemes in relation to a corresponding continuous speech signal. An accurate and fast alignment procedure is a necessary tool for developing speech recognition and text-to-speech systems, but also important tool for phoneticians and linguists in automatic phonetic analysis of speech. We used the algorithm described in [11] and its improvement in [12]. The algorithm is aimed at minimizing the epsilon-insensitive loss in expectation, and produces the best known results on the TIMIT dataset for American English [12]. The set of feature functions we used here is the same set used in [11], except for replacing the frame-based classifier with the one described in Section 3.2. We build a new model for durations, and remove the rate feature function. For test purposes, we use the similar settings as we did for the phoneme classifier. Results are listed in Table 1. Each row in the table means that if we allow, for example, the predicted boundary to be within 20 msec or less from the manually labeled boundary, then we get accuracy of 66.1% in the prediction.

3.4. Cantilation Signs Automatic Extraction

The result of the the alignment algorithm is a metadata file, which contains the exact boundaries of each phoneme as well as the location of each cantillation sign in the audio. With this information we segment the recording into words, marking the start time and end time of each word. Since in the Torah, each word can contain only one accent, word boundaries can also be used as the cantillation signs boundaries. We consider these word boundaries as 'high-level' boundaries for the cantillation signs, because the realization of the cantillation signs is effected by the number of syllables in the word.

Error Boundary	Success Rate
60 msec	85.9%
50 msec	83.2%
40 msec	79.2%
30 msec	73.3%
20 msec	66.1%
10 msec	50.9%

Table 1: Success rate of the phoneme alignment algorithm

3.5. Intonation analysis - 'finalis' note detection

One of the cantillation signs is called *Sof-Pasuk* or 'end of verse'. It is located under the last syllable of the last word in a verse and it marks the end of verse. In terms of the text content, each verse can be viewed as a simple or complex sentence, which stands as a closed unit. This fact is stressed in the Jewish Yemenite tradition due to the fact that during the liturgical event, in which the Torah is read, after each verse the reader pauses and another reader reads the Aramaic translation of the verse.

We extract the pitch contour of each cantillation sign using the SWIPE pitch detection algorithm [13] with non-overlapping window of size 10ms. Figures 2 and 3 shows a typical performance realization from two readers of the Sof-Pasuk sign.

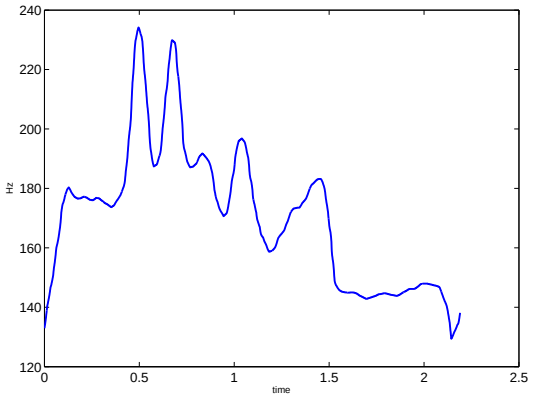

Figure 2: A typical Sof Pasuq (end of verse) pitch contour (reader A)

Listening analysis of a collection of Sof-Pasuk realizations together with viewing their pitch contour graphs reveals an interesting and quite stable structure. The melodic curve contains a melisma with a number of ornamental pitches before the reader arrives to the ending note. We refer to this note as the finalis note. This note is characterized by a reltovely stable pitch and a relatively long duration. An ethnomusicologist who is familiar with the Yemenite Jews tradition and he has confirmed that the musical realization of the Sof-Pasuk sign always ends on the finalis note.

We adopted the analysis method as was done in [6] and used the kernel density estimation of the pitch curve to estimate its distribution. The pitch density estimation on one realization of Sof Pasuk is depicted in Figure 4. We chose the peak after disregarding all pitches with less than 5% of the density. If more than one peak exists we took the one with the lowest pitch.

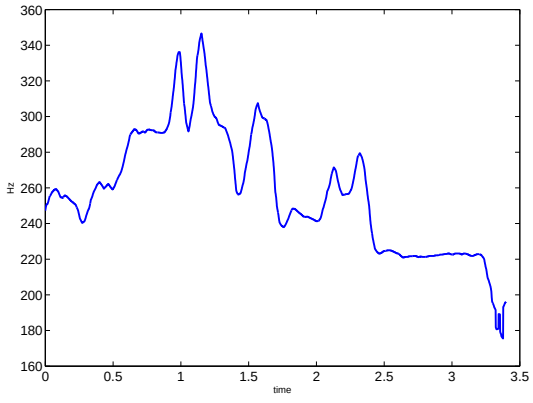

Figure 3: A typical Sof Pasuq (end of verse) pitch contour (reader B)

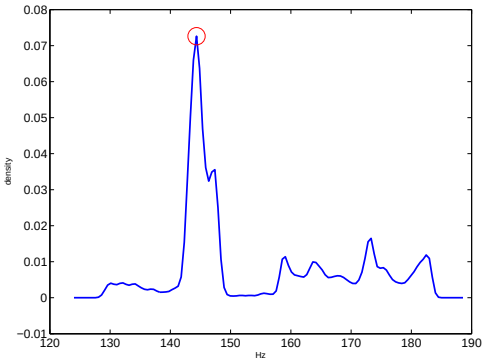

Figure 4: Pitch distribution of Sof-Pasuk and detection of finals note

4. Results

Using the methods described above, we analyzed two chapters of reading the Torah in the Jewish Yemenite tradition. In the following we refer to Rabbi Shlomo Kare as Reader A and Rabbi Shalom Cohen as reader B. The analysis steps included: (i) split the chapter into verses using end of verse symbol; (ii) convert each verse vocalized text to a sequence of rich phonemes file, which contains both the phonemes as well as the tropes exact location; (iii) manually split the recording of the chapter into verses; (iv) apply the alignment algorithm to align the phonemes and the cantillation signs to the audio; (v) extract all the realizations of Sof Pasuk; (vi) compute pitch contour of each Sof Pasuk instance; and (vi) extract the finals note of each Sof Pasuk.

In the first example, both readers read the first chapter of the book of Genesis. Figure 5 shows the readers finalis note (in Hertz) along this unit. We can see that within the chapter the finals note is not stable. There are ”cycles” along the chapter such that in each cycle the reader?s finals note starts on a relatively low pitch and gradually shifts upward up to a point where the reader decides to start a new cycle. To decide whether the reader had started a new cycle or whether he fluctuates inside a cycle we computes the standard deviation of pitch difference between all the verses in each chapter. We than marked as new cycles only verses where the difference between a given verse and a previous mark is greater than one standard deviation. These are marked as new cycles by the blue and black circles in Figure 5 and Figure 6.

The red dashed lines in Figure 5 show the division of this chapter to small contextual units, arranged by their subject. This division was done in a special edition of the bible, called ’Da’at Mikra’ which has a comprehensive commentary on the biblical text. The division of this first chapter is straightforward because it is divided into six units, where each unit tells the story of the creation of a single day from the first day to the sixth day. The text of the last day of creation, is in the first three verses of the second chapter and we analyzed all of them as one unit. From the graph we see that the division of reader B is quite consistent with the contextual division of the chapter. For reader A it is less consistent but there are two cycles which do overlap.

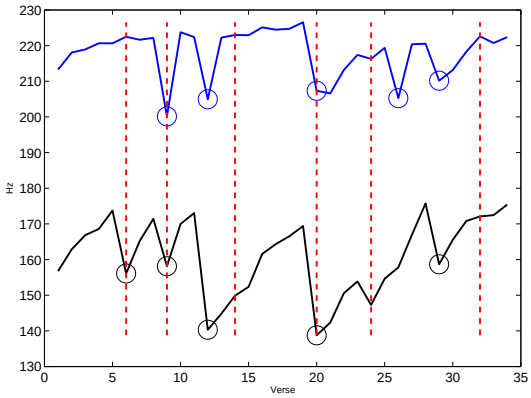

Figure 5: The finals of Sof Pasuq of reader A (blue) and reader B (black) of Chapter 1 of Genesis. The dashed red lines indicate the beginning of a section or subsection in the chapter. The black and blue circles are automatic units selection

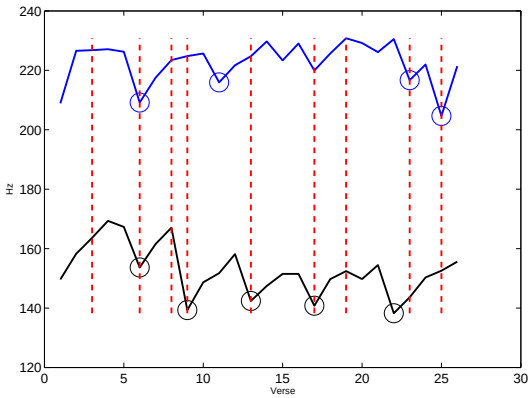

Figure 6: The finals of Sof Pasuq of reader A (blue) and reader B (black) of Chapter 4 of Genesis. The dashed red lines indicate the beginning of a section or subsection in the chapter. The black and blue circles are automatic units selection

The second example is Chapter 4 of Genesis, which tells the story of Cain and Able. Figure 6 shows the pitch of the finalis along this chapter. Here, as in Figure 5 the dashed vertical red lines mark the division of the text into small units as was done by ’Daat Mikra’ bible. Although the division of this chapter into small units is not as straightforward as the first chapter of Genesis, we can still see from the graph that for both readers there are cycles that overlap with the contextual division of the chapter. For example, verse 6 starts a conversation between God and Cain. Both readers starts a cycle in this verse.

Table 2 and Table 3 show analysis of the pitch deviations of the readers across the units and at the beginning of units. We can see that Reader A lowers his finals pitch in an average of 120 cents at the beginning of each unit while reader B lowers the pitch in 217 cents. The increase in pitch inside the units is much more moderate - average of 26 cents to reader A and average of 51 to reader B.

	Genesis Chapter 1	Genesis Chapter 4	All
Reader A	-134	-105	-120
Reader B	-241	-192	-217

Table 2: Mean pitch shift in beginning of a unit (in cents)

	Genesis Chapter 1	Genesis Chapter 4	All
Reader A	27	26	26
Reader B	50	52	51

Table 3: Mean pitch shift inside units (in cents)

5. Discussion

The phenomenon described here of division of the text into subsections of a few verses each, extends the complex syntactic divisions at the verse level to the chapter level. Note that the system itself does not provide means at this level, and it is the reader who should, by means of performance communicate this information. The interesting phenomenon however is that this task is achieved through pitch variations of the finalis note rather than by other, simpler cues such as longer pauses. This demands an exceptional memory for pitch on the part of the reader, as well as presumably by the audience, especially given that between verses another reader provides the verse in Aramaic. This performance practice may also explain why the readers use pitch variations rather than longer pauses. Timing is a problematic cue since it is not controlled by the reader, rather it results from the interaction between two different readers. Hence the solution offered by creating cycles of rising and then falling pitch of the finalis is truely a remarkable one. This phenomenon could only be revealed by the automatic analysis following the text alignement procedure.

We plan as our next step to use this tool to collect more data from the yemenite tradition and to further analyze this performance practice. We claim that much of the variability in the performance is due to the requirement to perform the same melody motive on words that can have different syllable structure. Using the tools we developed, we can analyze the performance of the cantilation in relation to the syllables content of the words.

The tools that we have developed in this work can be used as foundation for a comprehensive study of the performance practice of jewish cantilation in different communities and the relation between them. In order to deploy these tools to other jewish communities we need to adapt our phoneme classifier and phoneme aligner tools that were trained to the hebrew dialect of yemenite jewish community to the hebrew dialect of other jewish communities.

References

[1] I. Yeivin: *Introduction to the Tiberian Masorah*. Published by Scholars Press for the Society of Biblical Literature and the International Organization for Masoretic Studies, Missoula, Mont., 1980.

[2] J. D. Price: *The syntax of masoretic accents in the Hebrew Bible*. E. Mellen Press, Lewiston, N.Y., 1990.

[3] A. Z. Idelsohn: *Jewish music in its historical development,*. Schocken Books, New York, 1967.

[4] U. Sharvit: *The musical realization of biblical cantillation symbols in the Jewish Yemenite tradition*. In *Yuval*, volume 4:179–209, 1982.

[5] A. Hanoch: *The Ashkenazi tradition of Biblical chant between 1500 and 1900 - documentation and musical analysis*. Tel-Aviv University, Faculty of Fine Arts, School of Jewish Studies, Tel-Aviv, 1978.

[6] P. Kranenburg, D. P. Biró, S. R. Ness, and G. Tzanetakis: *A computational investigation of melodic contour stability in Jewish Torah Trope Performance Traditions*. In *Proceedings of the International Society on Music Information Retrieval (ISMIR2011) Conference*. 2011.

[7] G. Dzhambazov, S. Şentürk, and X. Serra: *Automatic lyrics-to-audio alignment in classical Turkish music*. In *4th International Workshop on Folk Music Analysis*. Istanbul, Turkey, 2014.

[8] *http://www.mechon-mamre.org/i/t/t0.htm*.

[9] K. Crammer, O. Dekel, J. Keshet, S. Shalev-Shwartz, and Y. Singer: *Online Passive Aggressive Algorithms*. In *Journal of Machine Learning Research*, volume 7:551–585, 2006.

[10] J. Keshet, D. McAllester, and T. Hazan: *PAC-Bayesian Approach for Minimization of Phoneme Error Rate*. In *IEEE International Conference on Audio, Speech and Signal Processing (ICASSP)*. 2011.

[11] J. Keshet, S. Shalev-Shwartz, Y. Singer, and D. Chazan: *A Large Margin Algorithm for Speech and Audio Segmentation*. In *IEEE Trans. on Audio, Speech and Language Processing*, volume 15(8):2373–2382, 2007.

[12] D. McAllester, T. Hazan, and J. Keshet: *Direct Loss Minimization for Structured Prediction*. In *Advances in Neural Information Processing Systems (NIPS) 24*. 2010.

[13] A. Camacho and H. Y. Flory: *A sawtooth waveform inspired pitch estimator for speech and music*. In *The Journal of the Acoustical Society of America*, page 1652, 2008.