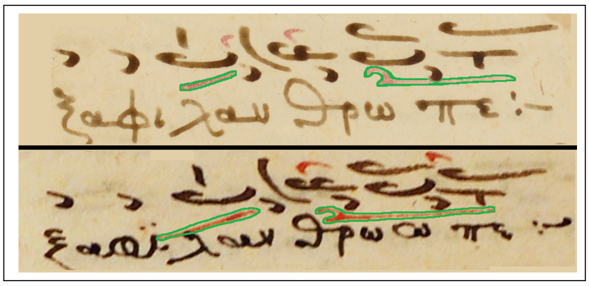

Figure 1. The top excerpt is from the Heirmologion of the manuscript Iveron 1167 f 4r, while the bottom is the same excerpt from the Heimrologion of the manuscript Iveron 1155 f 4r. As can be seen, the same Voiceless sign (red with a green border) spans several Voiced signs (black) in the two manuscripts.

biguous excerpts from corresponding manuscripts to locate a kind of pattern.

Thirdly, at a higher level, the MBn can be understood as a sequence of signs that indicate intervals. The pitches of these intervals are determined by a series of specific signs which are always placed at the beginning of the pieces. This group of signs is called Martyriai. Choosing a pitch as representative of each Martyria is not always straightforward [26]. The choices made in the specific music pieces are such as to facilitate the comparison of the three corpora.

Fourthly, our Knowledge Representation gives us the ability to capture the melodic phrases of each piece. The choice of end points of music phrases is not objective. In order to be consistent with our choices across the corpora, we have created a set of general rules and separation preferences.

Rule 1.1: Phrases are separated in the last syllable of a word. **Rule 1.2:** Phrases are separated in the last syllable of an *Enclisis* group (linguistic term). **Rule 2.1:** The last syllable of the phrase usually contains at least one of the following signs [3] : *Diple Apostrophos*, *Diple*, *Apoderma*, *Kratema* (rarer), *Klasma* (rarer). **Rule 2.2:** The penultimate syllable of the phrase can have the *Koufisma* sign. **Rule 2.3 (intuitive):** The last syllable of the phrase may have one of the following compound signs: *Omalon*, *Piasma*, and *Vareia*. **Rule 3:** There should be a colon and/or a comma in the lyrics of the music text. **Rule 4 (tendency):** Small phrases rather than large ones are preferred. **Rule 5:** The *Martyriai* appearing mid-text usually separate phrases.

3. METHODS OF COMPARISON

In order to draw an outline of the evolution of the Heirmologic subgenre of Byzantine music, we use three methods to capture the differences of some aspects of the corpora: Notational texture (method 1), similarity of the melodic arches (method 2), and distances of Markovian models of the attributes (method 3).

[3] These are the names of the signs of MBn that etymologically come from Greek words [7].

3.1 Method 1: notational texture.

In plainchant music (e.g., Byzantine, Gregorian, Mozarabic etc.) a characteristic of the style of a piece is its average number of notes, time beats, or signs per syllable [27–33]. Based on this characteristic, we can cluster the music pieces into three categories: syllabic, neumatic, and melismatic. In order to eliminate any ambiguities in these categories, we define two types of texture: Notational and Durational texture. *Notational texture* of a corpus is defined as the average number of pitches or of signs on a syllable, while *Durational texture* of a corpus is defined as the average number of durations per syllable. When the average number of the measured quantity (pitch, duration, signs etc.) of a corpus is between 1 and 2, then the texture of the corpus is characterized syllabic, when it is between 2 and 4 then the texture of the corpus is characterized neumatic, and when it is 4 or more then the texture of the corpus is characterized melismatic. Hence, a music corpus in terms of pitch can belong to syllabic style, while in terms of duration can belong to neumatic style. The interpretation of the MBn in terms of the duration of the syllables remains an open question. Consecutively, as the music pieces considered in this research are written in MBn, our measurements cover the notational texture.

All the pieces of our corpora belong to Heirmologic subgenre. This means that the general notational texture of these three corpora is syllabic. Nevertheless, through this measurement we can obtain a more subtle distinction of the notational texture of the syllables as examined by Troelsgard [30]. Specifically, we measure the number of pitches per syllable, $notational\_texture = \frac{|voiced\_units|}{|syllables|}$.

3.2 Method 2: similarity of the melodic arches

As we discussed in section 2, tradition plays an influential role in the composition of a hymn. In this research, we are interested in (a) identifying the areas where the new compositions follow the tradition, (b) the areas where the new compositions diverge from the tradition, and (c) quantify the divergence of the new and prior compositions.

Plotting the melodic lines of the different compositions of the same poetic verses lined up by syllable, we observe that the compositions usually present similar melodic shapes in corresponding areas (Figure 2). Based on this observation and inspired by the idea of melodic arches [1, 2], we create a novel method that, given two melodies of the same number of syllables, returns a percentage of similarity of their melodic arches. The design of this algorithm is driven by (a) the fact that the compositions in Byzantine Music are rooted on the syllables of the lyrics, (b) the duration of the pitches is unknown, and (c) the need for the derivation of a comparison metric instead of a metric for the identification of a dominant melodic arches [1].

The algorithm consists of three phases, (a) the plotting of the pitches (Algorithm 1), (b) the extraction of the arches (Algorithm 2), and (c) the comparison of the arches of the melodic lines (Algorithm 3).

Given two music pieces, the algorithm extracts the pitches, and plots them on a two dimensional grid (Algo-

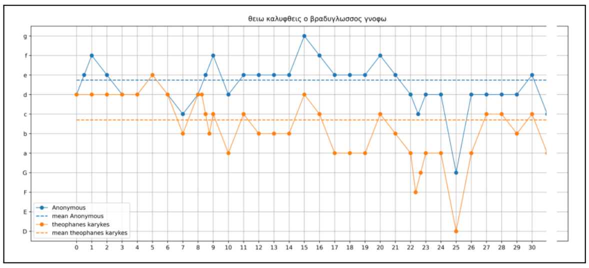

Figure 2. The melodic line of the first 30 syllables of Heirmos ‘Theio kalyptheis’ by the two compositions (Iveron 1101 and Iveron 1167). The x-axis represents the syllables, where at x=0, the value of y is the Martyria, i.e., the initial pitch (see [11]). The y-axis represents the steps (pitches).

rithm 1). The two pieces are aligned by syllables, and are of equal length. As a result, the pitches give us a view of the shape of the melodic lines, and highlights the similarities and differences of the compositions (Figure 2).

In the next phase the algorithm extracts the arches that are formed by the pitches (Algorithm 2), i.e. shape of the melody. Using these arches, we translate the two melodies into sequences of two types of arches: convex and concave.

In the final phase, the algorithm compares the compositions through their arches (Algorithm 3). When two areas are translated into the same arch type, they are considered to be similar. When two areas are translated into different arch types, they are considered to be dissimilar. When the areas cannot be translated into arches, they are classified as dissimilar due to different shapes.

When the arch of one melodic line corresponds to more than one arches of the other, the former is compared with the second degree polynomial regression of the latter. The second degree polynomial will give us a curve that is either convex or concave. Through the comparison of the two melodic lines, we obtain (a) the degree of similarity, (b) the degree of dissimilarity due to difference arches, and (c) the degree of dissimilarity due to difference shapes (Figure 3).

3.3 Method 3: distances of Markovian models of the attributes

As a third method, we capture the viewpoints as sequences (n-grams). We apply Markovian models on the n-grams to examine the divergence of the corpora. The melodic features of the notation are those that refer directly to the basic melodic line of the chant: pitch, interval, voiced unit. The information carried by the first two attributes is identical to that of the western staff. Voiced units are those that encapsulate the music information of the main melodic line as it is imprinted on the other two features (pitch, interval). In essence, the voiced unit attribute concerns the sign itself as it is imprinted in the score, while the pitch and the interval concern the aspects of the melody that the voiced units indirectly indicate (signified and signifier).

From the pitch and interval we also extract three others features which are redundancies of the melodic line: general syllabic pitch, general syllabic interval, and melodic

Algorithm 1 Pitch plotting algorithm.

1: Let $syllables$ be the sequence of syllables of a music piece.
2: Let $pitches$ be the sequence of pitches of a music piece.
3: Let $f^{assoc} : syllable_k \rightarrow pitches_k$ where $pitches_k$ is the sequence of pitches associated with $syllable_k$ and $pitches_{n_k}$ is the nth pitch of $pitches_k$.
4: Let $f^{pos} : syllable_{i=1}^{n} \rightarrow \mathbb{Z}^{+}$ be the mapping of the position of a syllable i in the sequence of the music piece to positive integers, this mapping defines the values of the x-axis.
5: $\forall n \in [\ 1,\ |pitches_k|\ ]$ where $pitches_k = f^{assoc}(syllable_k)$, we plot the points, $P = \{\ (x, y) \ : \ x = f^{pos}(syllable_{k-1}) + n/|pitches_k|,\ y = pitches_{n_k}\ \}$. Point (x_0, y_0) is reserved for the initial pitch given by a special sign known as Martyria.
6: Create an abstraction by removing the embellishments, in the context of syllable, from the melodic line, defined as $\mathbb{P}^{-} = \{(x, y) : (x, y) \in P, x \in \mathbb{Z}^{+}\}$, i.e. for every syllable keep the last pitch.
7: For each pair of points in $\mathbb{P}^{-}$, we identify the melodic contour (i.e., step up or down). Zero intervals inherit the melodic contour of the previous pair, hence they do not change the melodic contour.

contour of syllable. For the general syllabic pitch, instead of extracting every pitch from every sign of the syllable, we extract only the last pitch of the syllable. For the general syllabic interval, instead of extracting every interval from every sign of the syllable, we extract the sum of all intervals of the syllable. For the melodic contour of the syllable, we extract the values $(-, +, 0)$, which describe the contour of the general syllabic interval.

Using our Knowledge Representation [11], the corpora is translated into sequences of values based on the logic of viewpoints [20, 21]. For the aforementioned viewpoints, we train Markov models of different order, and we measure the distance of the models using the Jensen-Shannon distance (i.e., the square root of the Jensen-Shannon divergence) [34, 35]. To tackle the zero-probability problem we use Laplace smoothing [36].

Markov models express the probability of an n-gram appearing in the corpus. However, Markov models do not capture the position of the n-gram in the corpus and since we are restricted to low order Markov models due to overfitting, two different corpora can theoretically result in the same probability distribution. As we are using this method to find the relation of two corpora, we apply it both on the detailed dataset that reflects the music surface and on a simplified dataset that has a single value associated with each syllable. Since the corpora contain the same lyrics and in the simplified dataset every syllable has one value (one pitch, one interval, or one contour), the n-grams span the same regions for both melody and lyrics.

Furthermore, since the divergence is derived from the

Algorithm 2 Arch extraction algorithm.

Let $\alpha = (start, peak, end, shape)$ be an arch, where $start,\ peak,\ end \in \mathbb{P}^-$ and $shape \in \{CONVEX, CONCAVE\}$.
Let point x be the starting point, then $start^\alpha = x$.
Find x_i such that all pairs in the range $[x, x_i]$ have the same contour or zero intervals, then $peak^\alpha = x_i$.
Find x_j such that all pairs in the range $[x_i, x_j]$ have the same contour or zero intervals, then $end^\alpha = x_j$.
Set $shape^\alpha = CONCAVE$ when $start^\alpha < peak^\alpha$, otherwise $shape^\alpha = CONVEX$.
{Since zero intervals do not change the contour, the arch will include zero intervals.}

Markov tables, and since we are using it for comparison, we make sure that the Markov tables of the corpora under comparison consist of a common set of n-grams, spanning the union of the set of symbols found in the corpora under comparison. As such, the divergence reflects the distance of the probability distributions over a single set of n-grams.

We study n-grams between the orders of 2 and 4, and favour higher orders until the point where all lines show decrease due to over-fitting. As we are interested in the comparison of the corpora, higher orders with the same probability indicate areas of similarity between the corpora. However, with higher order Markov, we tend towards the scenario where an n-gram appears once in one corpus and not the other, the Markov table is dominated by noise due to smoothing, and Jensen-Shannon distance reflects the noise rather than the relation of the data.

Finally, the absolute values of the Jensen-Shannon distance are not meaningful without a reference. As we are studying the distances of corpora, we need a reference in order to quantify the magnitude of their distance. For this reason, we make use of a case study corpus consisting of 16 Heirmoi in first echos which have the property that the 8 pre-Karykis [4] and 8 Karykis [5] Heirmoi belong to the pre-Karykis' music style. Generally, Karykis changes the compositional style in first echos, but for some reason, he keeps the previous style (pre-Karykis) for only these 8 Heirmoi. This is confirmed by the similar cadential pitches and pitch profiles, the highest melodic arch similarity (similar arches 84.4%, dissimilar arches 12.9% and dissimilar shapes 2.7%), and low Jensen-shannon distance (gray line in Figures 4 & 5). For these reasons, we consider the distance of these case study corpora as a threshold for similarity.

4. RESULTS

Table 1 presents the notational textures (i.e. the number of voiced units per syllable) of the three corpora. The three corpora show a syllabic notational texture. We observe that Karykis tends even closer to compose Heirmos with absolute syllabic notational texture (i.e., near to 1).

[4] Manuscript Iveron 1101, folios 7v-9r
[5] Manuscript Iveron 1167, folios 138r-141v

Algorithm 3 Arch comparison algorithm.

Let A and A' be the sequences of arches of the two melodic lines under comparison.
while $A \neq \emptyset$ and $A' \neq \emptyset$ **do**
  Let α_1 be the first arch in A, and α'_1 be the first arch in A'.
  Let $L = \emptyset,\ R = \emptyset$.
  if $start^{\alpha_1} < start^{\alpha'_1}$ **then**
    $L = \{\alpha_1\}$.
    $R = \{\alpha'_n \in A' : end^{\alpha'_n} \leq end^{\alpha_1}\}$.
  else if $start^{\alpha_1} > start^{\alpha'_1}$ **then**
    $L = \{\alpha'_1\}$.
    $R = \{\alpha_n \in A : end^{\alpha_n} \leq end^{\alpha'_1}\}$.
  else
    $L = \{\alpha \in A : end^{\alpha} \leq \max\limits_{\alpha'' \in \{\alpha_1, \alpha'_1\}} end^{\alpha''}\}$.
    $R = \{\alpha' \in A' : end^{\alpha'} \leq \max\limits_{\alpha'' \in \{\alpha_1, \alpha'_1\}} end^{\alpha''}\}$.
  end if
  Let $area^L = \sum_{i=0}^{n} end^{\alpha_i} - start^{\alpha_i}$ where $\alpha \in L$. Equally we define $area^R$.
  if $area^L \neq area^R$ **then**
    Expand smallest area by adding padding, i.e. zero intervals.
  end if
  Let $quad(area^L)$ and $quad(area^R)$ be the two quadratic regressions using the points of the arches and the padding included in these areas. The two areas are considered similar if the curves have the same shape, i.e., both curves being convex or concave.
  Remove from A and A' the arches in L and R.
end while

corpora	pre-karykis	karykis	balasis
notational texture	1.22	1.13	1.23

Table 1. The notational texture of the corpora.

Figure 3 shows the results of the similarity of the melodic arches. The melodic arches of the Karykis-Balasis corpora show the greatest similarity (78.99%). In the second place we have the pre-Karykis – Karykis pair with similarity (68.18%), and in the last place the pre-Karykis – Balasis pair with similarity (62.34%). The exact opposite is true for dissimilarity and the percentage of dissimilar shapes. Our results follow the observation of Stathis [13, pp. 26], quoting "the seventeenth-century Heirmologia contain kallopismoi [literally, to make beautiful or embellishments, but not to be confused with improvisations] of the Heirmologia of Theophanes Karykes and Iohasaph the New Koukouzeles", i.e. the 17th onwards Heirmologions are not considered original compositions as they use as a basis the 16th century Heirmologions.

Figures 4 & 5 present the Jensen-Shannon distances of the corpora on the studied viewpoints. We observe that Karykis-Balasis show the smallest difference, while pre-Karykis – Karykis and pre-Karykis – Balasis show similar

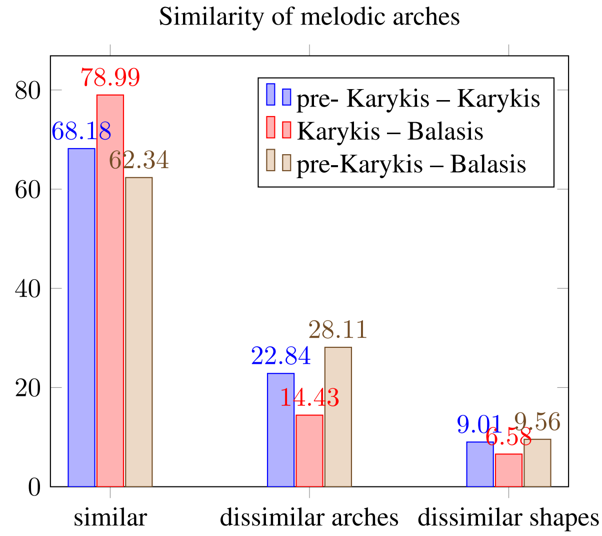

Figure 3. The bar charts of the melodic arches' similarity.

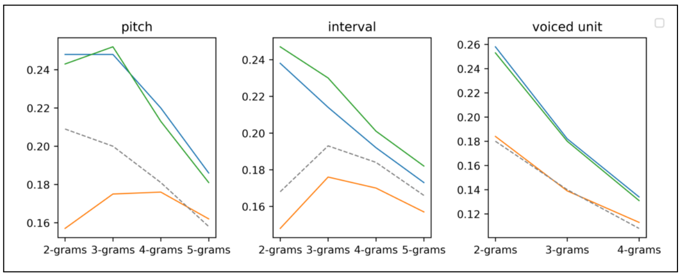

Figure 4. The Jensen-Shannon distances (abbr. JSd) of the Markov models. The reference line shows the JSd of the case study corpora (pre-Karykis and Karykis) where Karykis chooses exceptionally to keep the pre-Karykis tradition. The Blue line shows the JSd of the pre-Karykis – Karykis, the orange shows the JSd of the Karykis - Balasis, and the green line shows the JSd of the pre-Karykis – Balasis. The orders of n-grams are kept low in order to reduce the negative effect of higher order n-grams [38].

distances which are greater than of Karykis – Balasis. The close relation of Karykis – Balasis is confirmed by the near placement of the orange and gray lines of the case study corpus (section 3.3). The low rate of dissimilarity of the Karykis – Balasis Markov models is consistent with the observation of Makris [37] that the modality of Karykis – Balasis remains the same in contrast to the modality of the pre-Karykis era.

5. DISCUSSION

The notational texture (i.e. the number of voiced units per syllable) of the corpora outlines a change that occurred in the Heirmologic style. As shown in Table 1, the notational texture of Karykis presents a drop ($1.22 \rightarrow 1.13$) which leads us to a more syllabic texture. At the same time, however, Balasis shows an increase of notational texture, which reaches the same value as pre-Karykis ($1.13 \rightarrow 1.23$). If we consider that there is no important difference in the

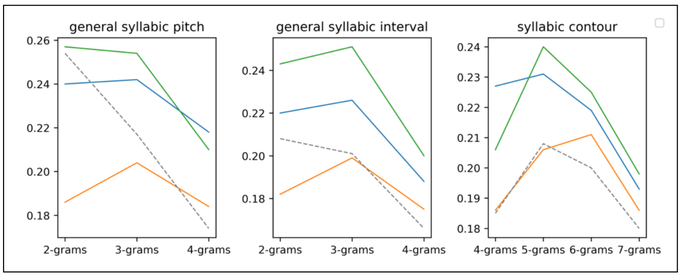

Figure 5. As of Figure 4, the Jensen-Shannon distances of redundant viewpoints.

interpretation-performance of the notation of the three corpora, then we can conclude that the values of the notation texture also correspond to values of durational texture. From these changes to the values of the notational texture, we can observe that Karykis simplifies the pre-Karykis melos and Balasis makes it sophisticated again by increasing the textures (notational and durational).

According to the bar chart of similarity of the melodic arches, the Karykis and Balasis pair present the greatest similarity than the other two. This presupposes that Karykis changes the tradition of the Heirmologion and Balasis largely follows Karykis. Also, we observe that all pairs present a melodic similarity that is more than 60%. This can be justified by the fact that the compared pieces are different compositions of the same lyrics (different melodies of the same Heirmoi). It seems that there are some constraints between the text and the melody, especially in the accented syllables. For example, we can guess that the composers want the accented syllables to be in local maxima. So, the peak areas influence the melodic contour of the melodic lines. This presupposes that Karykis changes the tradition and Balasis largely follows Karykis. Specifically, Karykis – Balasis have less distances than pre-Karykis – Karykis and pre-Karykis – Balasis ones.

These results can be viewed within the context of the *Theses* concept in Byzantine music. *Theses* are *formulae* that work as building blocks in Byzantine music pieces. Since the Markov models capture the melodies as frequencies of small patterns, we can infer that similar models of two corpora indicate that the two corpora use similar *Theses*.

6. CONCLUSIONS

This paper presented three related corpora belonging to the Heirmologic subgenre of the Byzantine music. The corpora are digitised using the Knowledge Representation of [11] which captures the MBn. Through this effort we envisage the creation of a large database of MBn scores. Through the computational analysis we observed that Karykis changes the style of the subgenre while the succeeding tradition of the subgenre (Balasis) follows the Karykis style. Even though the analysis presented addresses a specific research question, it offers us insights which will become the inspiration for further research in the domain of Computational Byzantine Musicology.

7. ACKNOWLEDGMENTS

We would like to thank Fabian Moss for his insights in the digitisation of the corpora.

8. REFERENCES

[1] D. Huron, "The Melodic Arch in Western Folksongs," *Computing in Musicology*, vol. 10, Mar. 1996.

[2] N. Cook, *A guide to musical analysis*, reprint ed. Oxford: Oxford Univ. Press, 2009.

[3] "Convention for the Safeguarding of the Intangible Cultural Heritage Intergovernmental Committee for the Safeguarding of the Intangible Cultural Heritage Fourteenth Session Bogotá," Dec. 2019.

[4] G. Stathis, *Ta protographa tes exegeseos eis ten nean methodou semiographias: ta prolegomena*. Athens, Greece: Institute of Byzantine Musicology, 2016, vol. 1.

[5] M. Velimirović, *Byzantine elements in early Slavic chant: The Hirmologion*. Munksgaard, 1960.

[6] E. Wellesz, *Eastern Elements in Western Chant: Studies in the Early History of Ecclesiastical Music*. Byzantine Institute, 1947, google-Books-ID: uXzQvwEACAAJ.

[7] C. Troelsgård, *Byzantine neumes: a new introduction to the Middle Byzantine musical notation*, ser. Monumenta musicae Byzantinae. Subsidia. Copenhagen : Lancaster: Museum Tusculanum Press ; Gazelle [distributor], 2011, no. 9, oCLC: ocn750956349.

[8] M. Alexandru, *Palaiographia Byzantines Mousikes [Paleography of Byzantine Music]*, Apr. 2017. [Online]. Available: https://repository.kallipos.gr/handle/11419/6487

[9] P. Savage, "An overview of cross-cultural music corpus studies," in *Oxford Handbook of Music Corpus Studies*. New York: Oxford University Press. [Online]. Available: https://www.researchgate.net/publication/326143358_An_overview_of_cross-cultural_music_corpus_studies

[10] X. Serra, "Creating research corpora for the computational study of music: the case of the compmusic project," in *Audio engineering society conference: 53rd international conference: Semantic audio*. Audio Engineering Society, 2014.

[11] P. Polykarpidis, C. Anagnostopoulou, and D. Kalofonos, "Towards the digitisation of Middle Byzantine notation: a formal model and Knowledge Representation," in *Music Encoding Conference 2022*, Halifax, Canada, To be published.

[12] ——, "MBnHeirmologicCorpora," Jul. 2022, publisher: OSF. [Online]. Available: https://osf.io/f97a5/?view_only=131fae4a07a04792986a12c43327709a

[13] G. Stathis, *Introduction to Kalophony, the Byzantine Ars Nova: The Anagrammatismoi and Mathēmata of Byzantine Chant*. Peter Lang, 2014, google-Books-ID: gAvwoQEACAAJ.

[14] D. Conomos, *The treatise of Manuel Chrysaphes, the Lampadarios: On the theory of the art of Chanting and on certain erroneous views that some hold about it*, ser. Corpus Scriptorum de re Musica. Wien: Österreichische Akademie der Wissenschaften, 1985, no. II.

[15] E. Makris, "Adjustments of modality in the Postbyzantine Heirmologion," in *Tradition and Innovation in Late- and Postbyzantine Liturgical Chant. Acta of the Congress Held at Hernen Castle, the Netherlands, in April 2005*. Hernen Castle, Netherlands: Peeters, 2008, pp. 37–63.

[16] N. Bouka, "E paradose ton Eirmologikon Byzantinon melodion tou Bareos echou apo ton 10o aiona eos ton 16o aiona," PhD, Ionian University, Corfu, Greece, 2004.

[17] A. Chaldaeakes, *Ermeneia syntomos eis tin kath' emas mousikin Pachomiou monachou*, ser. Mnimeia Ellenorthodoxis Mousikes. Athens, Greece: Athos, 2015, vol. 5, no. 1.

[18] D. Balageorgos and F. Kritikou, *Ta cheirographa tes Byzantines Mousikes: Sina*. Athens, Greece: Institute of Byzantine Musicology, 2021, vol. 2.

[19] G. Zesimos, "Kosmas Iberites kai Makedon: Domestikos tes mones ton Iberon," PhD, National and Kapodistrian University of Athens, Athens, Greece, 2007.

[20] D. Conklin and C. Anagnostopoulou, "Representation and discovery of multiple viewpoint patterns," in *Proceedings of the International Computer Music Conference (ICMC 2001)*, Havana, Cuba, 2001, pp. 479–485.

[21] D. Conklin and I. Witten, "Multiple viewpoint systems for music prediction," *Journal of New Music Research*, vol. 24, no. 1, pp. 51–73, Mar. 1995, publisher: Routledge_eprint: https://doi.org/10.1080/09298219508570672. [Online]. Available: https://doi.org/10.1080/09298219508570672

[22] L. Tardo, *Hirmologium Cryptense*, ser. principale (Facsimilés). Rome, Italy: Monumenta Musicae Byzantinae, 1951, vol. 3, no. 1.

[23] D. Balageorgos and F. Kritikou, *Ta cheirographa tes Byzantines Mousikes: Sina*. Athens, Greece: Institute of Byzantine Musicology, 2008, vol. 1.

[24] G. Stathis, *Ta cheirographa tes Byzantines Mousikes: Agion Oros*. Athens, Greece: Institute of Byzantine Musicology, 2015, vol. 4, no. 1.

[25] D. Touliatos-Miles, *A Descriptive Catalogue of the Musical Manuscript Collection of the National Library of Greece: Byzantine Chant and Other Music Repertory Recovered*. London: Routledge, Oct. 2017.

[26] J. Raasted, *Intonation Formulas and Modal Signatures in Byzantine Musical Manuscripts*. Copenhagen: E. Munksgaard, 1966.

[27] "Syllabic style," 2001. [Online]. Available: https://doi.org/10.1093/gmo/9781561592630.article.27236

[28] "Neumatic style," 2001. [Online]. Available: https://doi.org/10.1093/gmo/9781561592630.article.19788

[29] "Melismatic style," 2001. [Online]. Available: https://doi.org/10.1093/gmo/9781561592630.article.18333

[30] C. Troelsgård, "Stylistic variation in the old Sticherarion," in *Theoria and Praxis of the Psaltic Art*. Athens, Greece: Institute of Byzantine Musicology, 2006, pp. 225–232.

[31] G. S. Papadopoulos and P. Polykarpidis, "Elements in the development of how the Great Doxologies were sung from the 18th to the 19th century," *Journal of the International Society for Orthodox Music*, vol. 4, no. 2, pp. 293–322, Nov. 2020, number: 2. [Online]. Available: https://journal.fi/jisocm/article/view/95235

[32] P. Polykarpidis and G. S. Papadopoulos, "Melismatic elements in heirmologic chants from the 16th to the 19th century." Thessaloniki, Greece: IMS Study Group, 2022. [Online]. Available: https://www.academia.edu/52588244/Melismatic_elements_in_heirmologic_chants_from_the_16th_to_the_19th_century

[33] D. Spanoudakis, "Octaechia and Melismaticity in Byzantine Music. Collations and Musicological Reductive – Statistical Analysis," *IMS Regional Association for the Study of Music of the Balkans*, pp. 181–199, 2021. [Online]. Available: https://ejournals.lib.auth.gr/smb/article/view/7939

[34] F. C. Moss, "Transitions of Tonality: A Model-Based Corpus Study," PhD, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 2019. [Online]. Available: https://infoscience.epfl.ch/record/273178

[35] "scipy.spatial.distance.jensenshannon — SciPy v1.7.1 Manual." [Online]. Available: https://docs.scipy.org/doc/scipy/reference/generated/scipy.spatial.distance.jensenshannon.html

[36] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge, UK: Cambridge University Press, 2008.

[37] E. Makris, "Adjustments of modality in the Postbyzantine Heirmologion," in *Tradition and Innovation in Late- and Postbyzantine Liturgical Chant. Acta of the Congress Held at Hernen Castle, the Netherlands, in April 2005*. Hernen Castle, Netherlands: Peeters, 2008, pp. 37–63.

[38] J. Sakellariou, F. Tria, V. Loreto, and F. Pachet, "Maximum entropy models capture melodic styles," *Scientific Reports*, vol. 7, no. 1, Aug. 2017. [Online]. Available: https://www.nature.com/articles/s41598-017-08028-4

PLAYING TECHNIQUE DETECTION BY FUSING NOTE ONSET INFORMATION IN GUZHENG PERFORMANCE

Dichucheng Li[1] **Yulun Wu**[1] **Qinyu Li**[2] **Jiahao Zhao**[1]
Yi Yu[3] **Fan Xia**[2] **Wei Li**[1,4]

[1] School of Computer Science and Technology, Fudan University, Shanghai, China
[2] College of Experimental Art, Sichuan Conservatory of Music, Sichuan, China
[3] National Institute of Informatics (NII), Tokyo, Japan
[4] Shanghai Key Laboratory of Intelligent Information Processing, Fudan University, China

21210240219@m.fudan.edu.cn, {20110240018, 20210240306, weili-fudan}@fudan.edu.cn,
yiyu@nii.ac.jp, {935362804, 409769992}@qq.com

ABSTRACT

The Guzheng is a kind of traditional Chinese instruments with diverse playing techniques. Instrument playing techniques (IPT) play an important role in musical performance. However, most of the existing works for IPT detection show low efficiency for variable-length audio and provide no assurance in the generalization as they rely on a single sound bank for training and testing. In this study, we propose an end-to-end Guzheng playing technique detection system using Fully Convolutional Networks that can be applied to variable-length audio. Because each Guzheng playing technique is applied to a note, a dedicated onset detector is trained to divide an audio into several notes and its predictions are fused with frame-wise IPT predictions. During fusion, we add the IPT predictions frame by frame inside each note and get the IPT with the highest probability within each note as the final output of that note. We create a new dataset named GZ_IsoTech from multiple sound banks and real-world recordings for Guzheng performance analysis. Our approach achieves 87.97% in frame-level accuracy and 80.76% in note-level F1-score, outperforming existing works by a large margin, which indicates the effectiveness of our proposed method in IPT detection.

1. INTRODUCTION

The Guzheng (古筝), which is also known as the Chinese zither, is a plucked 21-string Chinese musical instrument existing for over 2,500 years [1]. Chinese traditional music attaches great importance to the melody, so a large number of playing techniques are used to enhance the vividness of Guzheng performance. The pitch variation produced by pressing the strings with the left hand is even regarded as "the soul of Guzheng music" [2]. Instrument playing technique (IPT) detection aims to classify the types of IPTs and locate the associated IPT boundaries in an audio clip. However, there were only few researches about IPTs in the field of Music Information Retrieval (MIR). Particularly, most researches on automatic music transcription (AMT) only consider pitch estimation, while a complete transcription of a Guzheng performance should contain the notations of the playing techniques as shown by the Guzheng numbered musical notation [1] in Figure 1. In this work, we propose an IPT detection method that can be further incorporated into a full transcription system of Guzheng music.

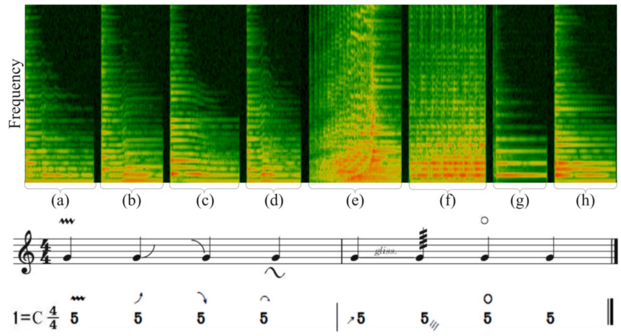

Figure 1. The spectrogram, staff notation and numbered musical notation of a Guzheng phrase that contains 8 notes with different playing techniques: vibrato (a), upward portamento (b), downward portamento (c), returning portamento (d), glissando (e), tremolo (f), harmonic (g), plucks (h)

One major difficulty for the IPT research is the lack of IPT sound databases. Most of the researches [3–6] on IPT recognition limited their experiments to samples from isolated notes in a single sound bank. However, according to [7], the performance of classifiers trained and tested with a single sound bank provides no assurance in the generalization capabilities of the classifiers. Ducher et al. [8] extended the above theory to IPT recognition research. They

[1] a musical notation system widely used in China and called *Jianpu* in Chinese

used five available IPT sound banks in their experiments but did not annotate any recorded corpora of audio.

Existing methods for IPT detection are mainly implemented through two steps [5,6,9]: locating and classifying. These methods are sensitive to the errors caused in locating. Recently, some end-to-end methods [10] have been proposed, while they are weak for variable-length audio and show low accuracy at the boundary of adjacent notes.

Note onset detection aims to localize the very beginning of each note. The beginning of a Guzheng note is easier to identify because the amplitude of that note is at its peak. In Guzheng performance, strings are usually plucked using special nails attached to fingers bottoms. When string is plucked, it has a pre-attack before the string reaches its full level vibration so the onset has a unique broadband spectrum. As shown in Figure 1, each Guzheng playing technique is applied to a note and each note in Guzheng music starts with a clear onset. So onset information is crucial to the Guzheng playing technique detection. However, to the best of our knowledge, no research has taken use of the onset information in the end-to-end IPT detection field.

The main contributions of this paper are as follows: 1) We create a new dataset, GZ_IsoTech, which consists of Guzheng playing technique clips from two Guzheng sound banks and real-world recordings recorded by a professional Guzheng performer; 2) We propose the first end-to-end method that can be applied to variable-length audio for Guzheng playing technique detection; 3) We propose a decision fusion method where an onset detector is trained to divide an audio into several notes and the predicted IPTs produced by the IPT detector are added frame by frame inside each note to get the IPT class with the highest probability within each note as the final output of that note.

2. RELATED WORK

Although the research on IPT detection is still in its early stage, we can summarize relevant researches and divide the development of this field into three periods.

The researches in the first period mainly focus on the playing technique classification of isolated notes [4, 11]. However, in reality, there are often continuous notes with varying playing techniques in an audio. We not only need to classify the IPT types but also locate the IPT boundaries.

The methods for IPT detection proposed in the second period can be divided into two steps: locating and classifying. In [5,6], signal processing methods were used to select candidates of the playing techniques in the melody contours extracted from 42 electric guitar solo tracks. Then the timbre and pitch features of the candidates were input into the classifiers such as Support Vector Machine (SVM). The candidates were manually selected based on the melody contour, so the methods are sensitive to errors in melody extraction and can hardly be generalized to other IPTs.

In [9], the candidates of guqin playing techniques were firstly located according to the IPT onset and duration annotations in 39 guqin solo recordings and then classified into six left-hand IPT types. Yet, we generally do not have onset and note duration as prior information in reality.

In [12], two binary classifiers bases on Convolutional Neural Networks (CNN) are firstly used to decide whether a fixed-length portion in a piano recording is played with the sustain pedal. Then sliding windows are used to apply the method to pieces with variable lengths. In [8], five different methods were proposed to classify 18 different IPTs in the cello solo audio concatenated by isolated cello notes with different IPTs from 5 different sound banks. Although the models proposed in [8, 12] can detect the IPTs in audio with continuous IPTs, the essence of them is to classify IPTs in a single block. These methods have redundancy in the computation because the input sound is split into overlappping frames that are fed individually to the network through a sliding window. Each convolution is thus computed several times on the overlapping parts.

We regard end-to-end IPT detection as the approach in the third period. In [10], the authors presented an end-to-end method based on Fully Convolutional Networks (FCN) to detect the IPTs in 10-second segments concatenated by isolated erhu notes. However, the method shows low accuracy at the boundary of adjacent notes and has computational redundancy when applied to variable-length audio.

Our model is built upon the FCN model presented in [10], with some improvements. We add a dedicated onset detector to take advantage of the significance of note onsets in Guzheng music. A decision fusion is implemented for the onset and IPT prediction to make a note-level prediction. In this way, we achieve a better performance for IPT predictions at the note boundaries. To apply our method to audio with variable lengths, 1D max-pooling layers which only halve the length of frequency axis are used.

3. GZ_ISOTECH DATASET

In this section, we introduce the Guzheng playing techniques considered in our work and the process of making the dataset.

3.1 Guzheng playing techniques

The traditional Guzheng playing techniques can be divided into two classes: plucking string with the right hand and bending string with the left hand [13]. In our work, we consider the following eight playing techniques which are the most frequently used in Guzheng solo compositions.

Left-hand playing techniques:

- **Vibrato (chanyin颤音):** the periodic oscillation of tones caused by the periodic pressing of a string with left hand (region (a) of Figure 1).

- **Upward Portamento (shanghuayin上滑音) (*UP* for short thereinafter):** press a string with left hand to increase the pitch of a ringing note to a desired pitch within major third (region (b) of Figure 1).

- **Downward Portamento (xiahuayin下滑音) (*DP* for short thereinafter):** decrease the pitch of a ringing note by releasing a bended string (region (c) of Figure 1). It is the opposite of Upward Portamento.