

A DATASET FOR GREEK TRADITIONAL AND FOLK MUSIC: Lyra

Charilaos Papaioannou¹, Ioannis Valiantzas², Theodoros Giannakopoulos³,
Maximos Kaliakatsos-Papakostas⁴, Alexandros Potamianos¹

¹ School of ECE, National Technical University of Athens, Greece

² Department of Music Studies, National and Kapodistrian University Of Athens, Greece

³ National Center for Scientific Research - Demokritos, Greece

⁴ Athena RC, Greece

cpapaioan@mail.ntua.gr

ABSTRACT

Studying under-represented music traditions under the MIR scope is crucial, not only for developing novel analysis tools, but also for unveiling musical functions that might prove useful in studying world musics. This paper presents a dataset for Greek Traditional and Folk music that includes 1570 pieces, summing in around 80 hours of data. The dataset incorporates YouTube timestamped links for retrieving audio and video, along with rich metadata information with regards to instrumentation, geography and genre, among others. The content has been collected from a Greek documentary series that is available online, where academics present music traditions of Greece with live music and dance performance during the show, along with discussions about social, cultural and musicological aspects of the presented music. Therefore, this procedure has resulted in a significant wealth of descriptions regarding a variety of aspects, such as musical genre, places of origin and musical instruments. In addition, the audio recordings were performed under strict production-level specifications, in terms of recording equipment, leading to very clean and homogeneous audio content. In this work, apart from presenting the dataset in detail, we propose a baseline deep-learning classification approach to recognize the involved musicological attributes. The dataset, the baseline classification methods and the models are provided in public repositories. Future directions for further refining the dataset are also discussed.

1. INTRODUCTION

Traditional music of Greece is under-represented in available datasets, despite the fact that this music offers unique perspectives that combine characteristics of Western and Eastern music. The development of a traditional Greek

music dataset is interesting to study in its own right, while it is also expected to provide a more complete picture of the music in the Mediterranean and Europe. Computational methods have been extensively studied in the past decades for carrying out ethnomusicological studies, constituting the field of Computational Ethnomusicology. A review of such methods is presented in [1], where it is evident that there are many benefits in compiling datasets of traditional music that can readily be used in computational models.

Several such datasets have been developed. For instance, an enormous database of Dutch melodies and songs that allows studying multiple musicological aspects is presented in [2]. Similarly, a dataset of Indian art music was presented in [3], where various MIR-related tasks were adjusted and applied therein. The employment of standard MIR tools has been evaluated in Arab-Andalusian music [4] and Flamenco music with the COFLA dataset [5]. Iranian Dastgah classification has been studied with MIR tools in [6]. A dataset of Georgian vocal music from historic tape recordings of three-voice songs was presented in [7], where the aim was to preserve cultural heritage and also to apply and examine existing MIR methods (e.g., for pitch and onset detection) in data of this form. Such datasets also provide opportunities for studying idiosyncratic musical instruments, e.g., in the case of Chinese music [8], where many instruments are employed that are not related to the ones used in Western music.

In several cases, off-the-shelf MIR tools are not well-suited for tasks that include traditional musical types. For instance, studying melodic similarity of Turkish non equally-tempered makam phrases in symbolic format required the development of a novel representation based on MIDI [9]. Similarly, rhythmic attributes of makam music with new models that integrate note transition rules for this music had to be developed for lyrics-to-audio alignment [10]. In addition, a novel method, based on audio signal processing, was created for identifying asymmetric rhythms in Greek traditional music [11].

The “Lyra” dataset of traditional and folk Greek music aims to contribute to all the aforementioned fields, from data-driven (computational) ethnomusicology to shaping new directions of research for MIR tools. The majority of available datasets that concern Greek music are unsuitable



© C. Papaioannou, I. Valiantzas, T. Giannakopoulos, M. Kaliakatsos-Papakostas and A. Potamianos. Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). **Attribution:** C. Papaioannou, I. Valiantzas, T. Giannakopoulos, M. Kaliakatsos-Papakostas and A. Potamianos, “A Dataset for Greek Traditional and Folk Music: Lyra”, in *Proc. of the 23rd Int. Society for Music Information Retrieval Conf.*, Bengaluru, India, 2022.

for computational analysis due to their unstructured nature. An exception is the Greek Audio Dataset (GAD) [12], which includes 1000 pieces with audio content (YouTube links), lyrics and annotations for genre (coarse categories including traditional, pop, rock), mood (valence-arousal plane coordinates) and pre-computed audio features; this dataset was later expanded in the Greek Music Dataset (GMD) [13], consisting of 1400 audio pieces with the aforementioned data. However both the GMD and the GAD are not focused on traditional and folk music, while the quality of audio recordings is varying significantly between genres.

The Lyra dataset presented in this paper, in contrast to the GMD, includes mainly traditional and folk Greek music with fine-grained labeling, focusing on musicological aspects of interest. Additionally, the recording quality is homogeneous across all pieces. Musicological soundness and high quality content are ensured by the fact that data has been collected and annotated from a documentary series that was presented by academics in Greek television. Some baseline classification tasks that are of particular interest in this dataset are presented, namely classification of pieces in genres, instruments and places of origin. In all cases, the results show that computational analysis can provide useful insight about musicological relations and phenomena in traditional and folk music of Greece – and, possibly, of other places.

2. DATASET EXTRACTION AND DESCRIPTION

2.1 Challenges and Methods

Large amounts of clean data is fundamental for current AI models to achieve their full potential. In this paper, we walked all the way, from the “data in-the-wild” multimedia content of a TV show to a fully annotated dataset, through a combination of machine automation and human evaluation/annotation processes. The consistency of the dataset and the richness of information it provides are tested by developing and training models that perform three different classification tasks.

In the case of Greek traditional and folk music, there are few cases where metadata is combined with recordings in a structured manner. Additionally, there is a matter of quality of recordings as it is significantly affected by various factors, including the equipment used, the social occasion (e.g., during a festival or inside a studio) and the time period in which it took place, i.e., older recordings tend to be of lower quality.

An integration of dissimilar recordings, in terms of quality, can introduce significant deficiencies towards studying the musicological characteristics of world music with computational tools. In order to truncate the effect of the audio quality factor, we decided to incorporate the episodes from the Greek documentary series “To Alati tis Gis - Salt of the Earth” broadcasted by ERT (Hellenic Broadcasting Corporation), where primarily traditional and folk music is presented. The episodes were filmed during a 10 year period under strict production-level

specifications, resulting to very clean and homogeneous audio content while significant wealth of information is provided by the presenter and the guests in the form of narrations between music performances.

The presented dataset consists of both the multimedia content and the annotations of interest. The multimedia content is provided as start and end timestamps that correspond to a single music piece, as parts of a longer episode, which is available online. Regarding the annotations, a taxonomy of labels is defined, based on the potential purposes of studies that might involve this dataset, considering also what metadata information can be retrieved either directly from the source or be integrated by volunteer annotators during the data collection process.

The study of Greek traditional and folk music involves knowledge about (i) the instrumentation, (ii) the genres, (iii) the places of origin and (iv) the way listeners perceive this music in terms of “danceability”, among others. While musical instruments, genres and geography are semantically well-defined, the same can not be claimed for listeners’ perception. Having at hand the multimedia content, i.e., audio and video, can be helpful to this end. Annotation about whether a music piece is being danced during its live performance can reveal cultural characteristics regarding the way this piece is perceived by the community, because body movements play an important role in music perception [14].

As a result, the taxonomy consists of (i) the musical instruments participating in the performance of each music piece (singing voice is considered an instrument), (ii) the musical genres and sub-genres that are identified by musicologists in Greek music, (iii) the places of origin and (iv) whether the music piece is being danced during its performance.

Volunteer annotators, students of the Department of Music Studies, undertook the task of separating each episode in music pieces and also labeling each one of them according to the specified taxonomy. A helper website was utilized where the respective category labels were added. An account was created for each annotator for the label assignment task. Every piece was labeled by two annotators and the final labels are the set of them where both annotators agree. At the end, the dataset that contains the aforementioned annotations along with the timestamps and the respective video id for each music piece was extracted from the database of the helper site.

2.2 Dataset description

Lyra dataset is organized into a single table where each row corresponds to a music piece while the columns include the various metadata information. Table 1 demonstrates the metadata categories.

Beginning with the simplest metadata categories, in terms of description, “id” is a unique identifier for each piece, generated by its title, replacing Greek with Latin characters and spaces with dashes. As expected, the number of unique values will be the same with the number of pieces, namely 1570. The same stands for “start-ts” and

name	# unique	multi-label	description
id	1570	No	unique identifier of each music piece
instruments	32	Yes	instruments participating in performance
genres	32	Yes	music style annotations
place	81	Yes (to contain regions)	full hierarchy of the place of origin
coordinates	81	No	latitude and longitude
is-danced	2	No	binary value (0 or 1)
youtube-id	74	No	id of the episode available online
start-ts	1570	No	start timestamp of the piece
end-ts	1570	No	end timestamp of the piece

Table 1. Metadata contained in the dataset.

“end-ts” that denote the exact time (second) that a song starts and ends in the corresponding video. The duration of the music pieces sums to approximately 80 hours.

The column “youtube-id” contains the id under which the video of the full episode is available online. The count of unique values are essentially the number of episodes that were used for the creation of the dataset. A typical duration of an episode is roughly a hundred minutes. The “is-danced” binary label informs about whether a music piece is being danced by the guests of the show. The music pieces annotated with “1” are approximately 51% while in the rest of them no dance performance occurs.

The classification of Greek music in “genres” is a work that requires one to take into account a number of socio-cultural and anthropo-geographical criteria. At an abstract level, we can distinguish the music of urban centers in contrast to the music of rural areas of Greece, with the former including rebetiko, laiko, urban-folk among others, while the latter, the music of rural areas, is what we generally call traditional music. Figure 1 shows the frequencies of the genres in the dataset, with “traditional” being the dominant one constituting almost 78% of the total. Depending on the place of origin, the style of a traditional music piece varies accordingly and, thus, several sub-genres flourish, such as Epirotic for the songs originated from Epirus. The 32 unique values in this metadata category are separated into 5 distinct genres and 27 sub-genres.

From a musicological perspective, Greek traditional music can be divided into two large geographic areas, i.e., the island and the mainland Greece. Each one creates a

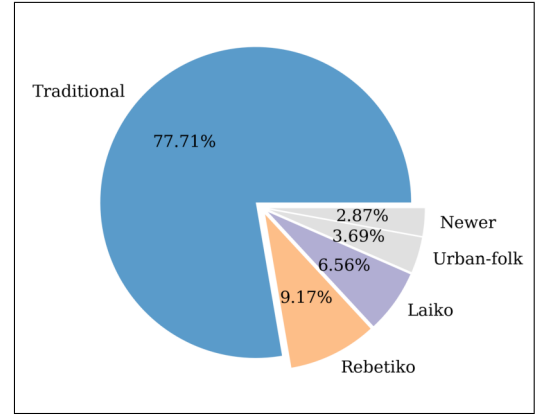


Figure 1. Relative frequencies of the music genres in the dataset.

distinctive musical feeling as there are large variations both on the rhythmic approach and the scales that are commonly used. For example, in islands we frequently come across music pieces with simple, fast rhythms while in the mainland more complex, slow rhythms are the norm.

The “place” (of origin) metadata category can be annotated with (i) a single label when the region from which a song derives is known, (ii) two labels when both region and a specific place are known and (iii) “None” denoting that there is not a specific place of origin for this piece. As an example, a music piece can be annotated with the region “Aegean sea” or with both “Aegean sea” and “Naxos”, an island of the Aegean sea, if this knowledge is available. Specifically, from the 81 unique places in the dataset, 20 are regions and only half of them include the remaining 61. The most represented regions can be seen in Figure 2.

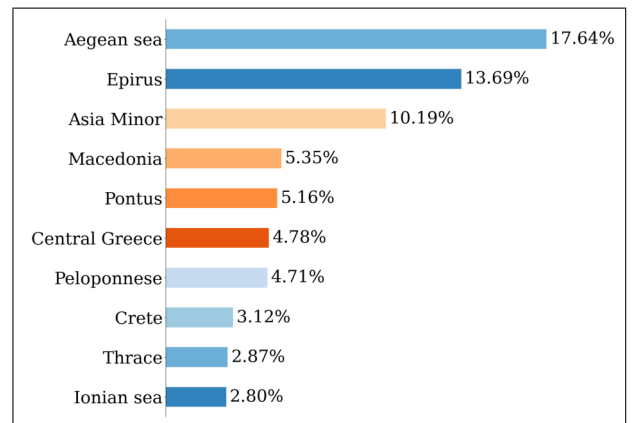


Figure 2. Relative frequencies of the most represented regions in the dataset.

The exact latitude and longitude of each place is also available at the “coordinates” column. The music pieces that do not have an explicit place of origin, such as the ones that belong to the “laiko” genre, are accounted for approximately 23% of the total. Figure 3 shows the location of the 81 places that exist in the dataset. We may notice the constant ability of music to exceed the borders; places where Greek culture thrived in the past and neighboring countries

that share the same tradition, form a mosaic of people that communicated freely with each other in a musical way that has reached towards us.



Figure 3. Map of all the places of the dataset.

Analogous connections, like the ones between genres and places, one expects to be observed between places and instruments as well. Indeed, for over 100 years, the established music ensembles of Greek traditional music are generally two, namely (i) those with the violin as leading instrument (often substituted by lyra and santouri), which have a greater presence in island Greece and (ii) those with the klarino (Greek clarinet) as the leading instrument, which is dominant in the mainland.

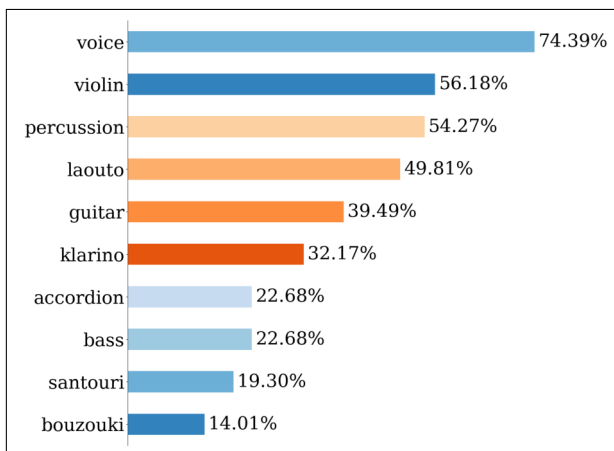


Figure 4. Relative frequencies of the most common musical instruments in the dataset.

In the popular and modern music domain, there is a great variety of instruments, but in most cases bouzouki, guitar, accordion and bass are common members of a laiko or rebetiko music ensemble. In the traditional music groups, the percussion and the laouto (Greek lute) are permanent companions of the leading instruments, offering melodic-harmonic background and rhythmic support. Of course, voice holds the main role in all kinds of performances.

In Figure 4 one can see the frequencies of the most popular instruments in the dataset. Singing voice is evident in almost 75% of it and instruments like violin, percussion and laouto, that have presence in both islands and mainland as well, are following.

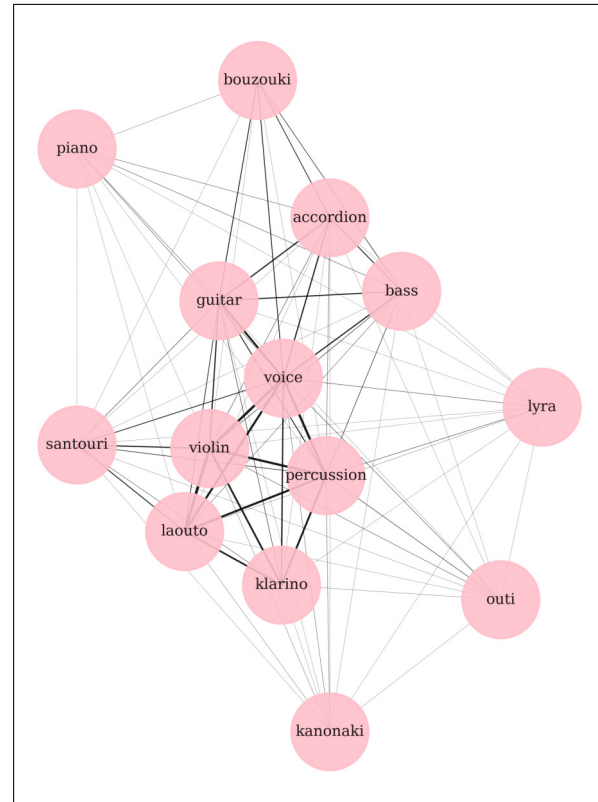


Figure 5. Graph that shows the co-occurrences of the fourteen most common musical instruments. The edge width is proportional to the number of music pieces a pair of instruments co-occur.

With regards to the music ensembles, it should be noted that 296 unique groups of instruments exist in the dataset, with the one constituted by voice, violin, percussion, laouto and klarino being by far the most popular by participating in the performance of around 12% of the music pieces. The co-occurrences of the most popular instruments can be seen in Figure 5 where the width of the graph edges is proportional to the number of pieces a pair of instruments co-occur in the dataset.

Sample rows of the dataset can be seen in Table 2. The dataset along with the baseline classification methods and the trained models are available online.¹

The shared metadata should be considered as the version 1.0 of the dataset. In the next versions, it will be evolved towards two main directions, namely (i) the incorporation of more metadata categories such as annotations according to the content of the lyrics, the lyrics themselves as well as information about the types of the dances that occur and (ii) the addition of more music pieces by following the same process either for next episodes of the same documentary series or for other series that have a similar theme.

3. BASELINE CLASSIFICATION

The audio recording of each music piece is represented using a Mel-scaled Spectrogram (mel-spectrogram): this is

¹ <https://github.com/pxaris/lyra-dataset>

id	instruments	genres	place	coordinates	is-danced	youtube-id	start-ts	end-ts
alexandra	voice violin percussion laouto klarino	Traditional Epirotic	Epirus Zagori	39.8648 20.9284	0	qrOwclmLFUk	749	927
choros-tik	percussion lyra	Traditional Pontian	Pontus	40.9883 39.7270	1	Aws0Y3aLaIs	1731	1886
agiothodo- ritissa	voice violin santouri percussion laouto guitar	Rebetiko	None	None	1	0cj8BNcAhg4	2632	2853
einai-arga- poly-arga	voice piano guitar bass bouzouki accordion	Laiko	None	None	0	zkoqg3VRVLA	2365	2614

Table 2. Sample rows of the dataset. Pipe “|” is used to separate values in a field.

a spectrogram whose frequencies are converted to the mel scale according to the equation:

$$m = 2595 \log_{10}(1 + \frac{f}{700}) = 1127 \ln(1 + \frac{f}{700}) \quad (1)$$

where **m** is the frequency in Mels and **f** is the frequency in Hz. Mel-spectrograms are calculated per fix-sized segment duration of 10 seconds. A non-overlapping window of 50 milliseconds has been applied, therefore the Mel-spectrogram size is 200 windows \times 128 frequency bins.

As a baseline classification approach, each 10-second Mel-spectrogram is classified to the aforementioned tasks (genre, place and instruments) using a Convolutional Neural Network (CNN). CNNs have been widely used in general audio [15], speech [16, 17] and music [18] classification tasks. In particular, we have adopted the following architecture: 4 convolutional layers of 5×5 kernels, single stride, and max pooling of size 2. The number of convolutional kernels (channels) are for the first layer 32, for the second 64, for the third 128 and for the fourth 256. The final output of the convolutional layers is passed through 3 fully connected layers, with the first having an output dimension of 1024, the second 256 and the third equal to the number of classes.

Note that fix-sized duration of segments is necessary, since audio recordings do not share the same size. Adopting a much longer segment would require zero padding for several mel-spectrograms and probably more CNN parameters. In addition, splitting the song into non-overlapping segments achieves some type of data augmentation. For two of the adopted classification tasks (namely genre and place), we have trained the CNNs using a multiclass, single-label setup, while for the instrument task, which is multi-label, we have trained multiple binary CNNs, one for each instrument, which have been evaluated separately.

After the training and validation procedure of each of the aforementioned CNNs, final testing was applied on the respective test recordings. For the test set, to avoid spreading pieces from the same broadcast across data splits, we separate training and test data on an episode level. From the 74 unique episodes, we randomly split 20% of them and use all the music pieces they include, namely 330, to form the test data. Obviously, this final testing needs to be carried out on a “song-level”, not a 10-sec segment level. Towards this end, a simple aggregation method was

adopted, by just averaging the posteriors of the individual segment decisions of the CNNs. This aggregated estimate was used as the final prediction and evaluated in the final testing.

4. RESULTS

The performance results during the training of the baseline classifiers are shown in Table 3, computed on a validation subset that corresponds to 20% of the segments. For the multi-label task (instrument recognition), we show the F1 metric for each binary subtask separately, while for the single-label tasks (genre and place) we show the overall macro F1 for all classes. We remind that this evaluation is performed on the validation split of the 10-second data.

Classifiers type	Task	F1 (%)
Multiclass classifiers	Genre	82.3
	Place	85.5
Binary classifiers for Instruments	voice	75.8
	violin	86.3
	percussion	97.1
	laouto	96.7
	guitar	80.2
	klarino	95.0

Table 3. F1 scores of the classifiers on the 10-second segments of the training data using a 20% validation subset.

As soon as the 10-second classifiers are trained, they are applied on the whole recordings of the testing data, and a simple majority aggregation is performed to extract the final decision, as described in the previous Section. For the instrument classification task, we compute the Area Under the Curve (AUC) metric per label (binary classification subtask). The results are shown in Table 4.

Finally, the confusion matrices along with the respective F1 measures for the multiclass, single-label classification tasks of “genres” and “places” are shown in Figures 6 and 7. All genres have been taken into account for the respective classification, but not the sub-genres. On “places” task, the pieces have been classified to the 10 most common regions (including “None”) plus the “other” category for the remaining. Genres classifier macro F1-score

Instrument	AUC (%)
voice	68.9
violin	85.2
percussion	95.1
laouto	93.8
guitar	73.5
klarino	90.9

Table 4. Area Under the Curve (AUC) scores of the instrument classifiers on the test data.

	Urban-folk	Laiko	Newer	Traditional	Rebetiko
Urban-folk	1	0	0	4	1
Laiko	0	1	0	1	6
Newer	0	0	0	3	0
Traditional	1	7	5	263	1
Rebetiko	0	0	0	13	23

Figure 6. Confusion matrix for “genre” classes on the test data. **Macro F1-score: 39.9%** and **Micro F1-score: 87.2%**.

is 39.9% and micro F1-score is 87.2%, while for the places classifier the macro F1-score is 34.4% and the micro F1-score is 42.4%.

5. DISCUSSION

A reason that the “voice” classifier has lower performance compared to the other ones may be of a musicological character. Indeed, while the presence of the rest of the instruments can depend significantly on the music style of a piece, the same does not apply to “voice” as it is the dominant musical instrument in any genre. Given the fact that the binary classifier is trained to recognize an instrument (evident in a part of a music piece) in each of the 10-second segments, regardless if it is present on it or not, we expect to move towards a space with latent musical features such as the music style, where “voice” may not be as discriminative as the rest of the instruments are.

With regards to confusion matrices, the misclassifications can be either due to imbalance between classes or statistical correlations across them. Specifically, for the “places” task, the confusions between regions that are geographically near may be justifiable, while for “genres” task the imbalance between the classes seems to have signifi-

	Aegean sea	Epirus	Crete	Macedonia	Asia Minor	None	other	Peloponnese	Pontus	Central Greece	Thrace
Aegean sea	20	2	0	0	7	0	0	0	0	1	0
Epirus	1	9	0	0	5	13	1	0	0	1	0
Crete	6	0	0	0	0	1	0	0	0	0	0
Macedonia	3	3	0	2	2	0	0	1	0	4	0
Asia Minor	17	2	0	0	12	5	1	0	0	1	0
None	4	2	0	0	1	51	0	1	0	1	1
other	10	2	0	1	16	30	6	0	0	2	0
Peloponnese	0	1	0	0	0	0	0	2	6	1	0
Pontus	2	0	1	1	0	0	0	0	25	0	0
Central Greece	0	1	0	0	8	2	0	0	2	8	0
Thrace	2	4	0	1	4	4	0	1	0	1	5

Figure 7. Confusion matrix for “place” classes on the test data. **Macro F1-score: 34.4%** and **Micro F1-score: 42.4%**.

cantly affected the performance of the model at the least represented ones.

We intend to improve models performance and conduct a more in-detail analysis in order to examine hypotheses such as the aforementioned ones in a follow-up study.

6. CONCLUSIONS

Greek traditional and folk music integrates components of Eastern and Western idioms, providing interesting research directions in the field of computational ethnomusicology. This paper presents “Lyra”, a dataset of 1570 traditional and folk Greek music pieces that includes audio and video (timestamps and links to YouTube videos), along with annotations that describe aspects of particular interest for this dataset, including instrumentation, geographic information and labels of genre and subgenre, among others. The advantage of this dataset is that the entire content is harvested from web resources of a Greek documentary series that was produced by academics with specialization in this music and, therefore, includes high-quality and rich annotations extracted from the content of the shows. Additionally, the production of recordings and video material is professional-level, providing a common ground in terms of audio quality. Three baseline audio-based classification tasks are performed, namely instrument identification, place of origin and genre classification.

The presented results indicate that specialized tasks, that use the audio signal, can potentially provide valuable insight about several aspects of this music. The combination of video and audio signals allows possible experimentation on methods that process multimodal data. The Lyra dataset includes material that readily allows MIR tools to be employed for reaching valuable musicological results. This dataset can also, potentially, foster the expansion of the MIR methods altogether.

7. ACKNOWLEDGEMENTS

The authors would like to thank Professor Lambros Liavas, the producer, researcher and presenter of the documentary series "To Alati tis Gis - Salt of the Earth". His work not only provides a fruitful source of knowledge and art but also acts as an inspiration for new researchers. This work would not have been possible without the help of the volunteer annotators. A big thank to the students of the Department of Music Studies of National and Kapodistrian University of Athens. Finally, we would like to thank the reviewers for their valuable and constructive comments.

8. REFERENCES

- [1] M. Panteli, E. Benetos, and S. Dixon, "A review of manual and computational approaches for the study of world music corpora," *Journal of New Music Research*, vol. 47, no. 2, pp. 176–189, 2018.
- [2] P. Van Kranenburg, M. De Bruin, and A. Volk, "Documenting a song culture: The dutch song database as a resource for musicological research," *International Journal on Digital Libraries*, vol. 20, no. 1, pp. 13–23, 2019.
- [3] A. Srinivasamurthy, G. K. Koduri, S. Gulati, V. Ishwar, and X. Serra, "Corpora for music information research in indian art music," in *Georgaki A, Kouroupetroglou G, eds. Proceedings of the 2014 International Computer Music Conference, ICMC/SMC; 2014 Sept 14-20; Athens, Greece.* Michigan Publishing, 2014.
- [4] R. C. Repetto, N. Pretto, A. Chaachoo, B. Bozkurt, and X. Serra, "An open corpus for the computational research of arab-andalusian music," in *Proceedings of the 5th International Conference on Digital Libraries for Musicology*, 2018, pp. 78–86.
- [5] N. Kroher, J.-M. Díaz-Báñez, J. Mora, and E. Gómez, "Corpus cofla: A research corpus for the computational study of flamenco music," *Journal on Computing and Cultural Heritage (JOCCH)*, vol. 9, no. 2, pp. 1–21, 2016.
- [6] S. Abdoli, "Iranian traditional music dastgah classification," in *Proceedings of the 12th International Society for Music Information Retrieval Conference*, 2011, pp. 275–280.
- [7] S. Rosenzweig, F. Scherbaum, D. Shugliashvili, V. Arifi-Müller, and M. Müller, "Erkomaishvili dataset: A curated corpus of traditional georgian vocal music for computational musicology," *Transactions of the International Society for Music Information Retrieval*, vol. 3, no. 1, 2020.
- [8] X. Gong, Y. Zhu, H. Zhu, and H. Wei, "Chmusic: A traditional chinese music dataset for evaluation of instrument recognition," in *2021 4th International Conference on Big Data Technologies*, 2021, pp. 184–189.
- [9] M. K. Karaosmanoğlu, B. Bozkurt, A. Holzapfel, and N. D. Dişiaçık, "A symbolic dataset of turkish makam music phrases," in *Proceedings of Fourth International Workshop on Folk Music Analysis (FMA2014)*, 2014, pp. 10–14.
- [10] G. B. Dzhambazov, A. Srinivasamurthy, S. Sentürk, and X. Serra, "On the use of note onsets for improved lyrics-to-audio alignment in turkish makam music," in *Devaney J, Mandel MI, Turnbull D, Tzanetakis G, editors. Proceedings of the 17th International Society for Music Information Retrieval Conference*, 2016.
- [11] T. Fouloulis, A. Pikrakis, and E. Cambouropoulos, "Traditional asymmetric rhythms: A refined model of meter induction based on asymmetric meter templates," in *Proceedings of the third international workshop on folk music analysis*, 2013, pp. 28–32.
- [12] D. Makris, K. L. Kermanidis, and I. Karydis, "The greek audio dataset," in *IFIP International Conference on Artificial Intelligence Applications and Innovations*. Springer, 2014, pp. 165–173.
- [13] D. Makris, I. Karydis, and S. Sioutas, "The greek music dataset," in *Proceedings of the 16th International Conference on Engineering Applications of Neural Networks (INNS)*, 2015, pp. 1–7.
- [14] M. Leman and P.-J. Maes, "The role of embodiment in the perception of music," *Empirical Musicology Review*, vol. 9, no. 3-4, pp. 236–246, 2015.
- [15] M. Papakostas and T. Giannakopoulos, "Speech-music discrimination using deep visual feature extractors," *Expert Systems with Applications*, vol. 114, pp. 334–344, 2018.
- [16] Z. Huang, M. Dong, Q. Mao, and Y. Zhan, "Speech emotion recognition using cnn," in *Proceedings of the 22nd ACM international conference on Multimedia*, 2014, pp. 801–804.
- [17] K. Choi, G. Fazekas, M. Sandler, and K. Cho, "Convolutional recurrent neural networks for music classification," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017, pp. 2392–2396.
- [18] M. Ashraf, G. Geng, X. Wang, F. Ahmad, and F. Abid, "A globally regularized joint neural architecture for music classification," *IEEE Access*, vol. 8, pp. 220 980–220 989, 2020.