

სტატიის ARTICLES

ლევან შულღიაშვილი

დამოუკიდებელი მკვლევარი, საქართველო
levanspublicemail@gmail.com

დავით შულღიაშვილი

თბილისის სახელმწიფო კონსერვატორია, საქართველო
david.shugliashvili@tsc.edu.ge

ნანა მჟავანაძე

ტრადიციული მრავალხმიანობის კვლევის საერთაშორისო ცენტრი
თბილისის სახელმწიფო კონსერვატორია, საქართველო
nana.mzavanadze @tsc.edu.ge

ფრანკ შერბაუმი

გეომეცნიერებათა ინსტიტუტი
პოტსდამის უნივერსიტეტი, გერმანია
fs@geo.uni-potsdam.de
DOI: 10.32859/kadmos/17/7-32

ქართული ტრადიციული მრავალხმიანობის ჰარმონიული ინტერვალური სტრუქტურა: საარქივო საველე ჩანაწერების (1948-1973) კომპიუტერული ანალიზის მონაცემები*

საკვანძო სიტყვები: გამოთვლითი ეთნომუსიკოლოგია, მრავალი
ფუნდამენტური სიხშირის ამოცნობა, ჰარმონიული ინტერვალები, ქა-
რთული მრავალხმიანობა, საარქივო აუდიოჩანაწერები

შესავალი

ქართული მრავალხმიანი სიმღერა ხასიათდება ხმების რთული კო-
ორდინაციითა და ვერტიკალური ორიენტაციით, რასაც მეცნიერები
საუკუნეზე მეტია იკვლევდნ და აღწერდნ (Аракишвили 1905, Nadel
1933; ახოზაძე 1957; ჭოხონელიძე 1988; Чхикваძე 1964; ასლანიშვილი

* კვლევა განხორციელდა შოთა რუსთაველის საქართველოს ეროვნული სამეცნი-
ერო ფონდის მხარდაჭერით (SRNSFG FR-21-2066).

1954 და სხვ.). დღემდე აქტუალურია დისკუსია რეგიონულ ტრადიციებში (სვანეთში, გურიაში, კახეთსა და სხვა კუთხეებში) ბგერათაწყობის, ინტერვალური აგებულებისა და მისი რაობის საკითხებთან დაკავშირებით (არაყიშვილი 1950; ერქვანიძე 1992, ვეშაპიძე, წერეთელი 2014; Mzhavanadze and Scherbaum 2020; Scherbaum, Frank, and Meinard Müller 2023; Scherbaum et al. 2022). ბოლო დრომდე, ჩანაწერების კორპუსის მასშტაბის აკუსტიკური ანალიზი ვერტიკალურ ინტერვალებზე ეყრდნობოდა ან ცალკეული ხმის F0 (ძირითადი სიხშირეების) ანოტაციებს (მაგ., არტემ ერქომაიშვილის ხმოვანი ჩანაწერები; Scherbaum, Müller, and Rosenzweig 2017), ან ჩაწერის სპეციალიზებულ მეთოდებს, როგორიცაა, მაგალითად, ხორხის მიკროფონები (Scherbaum 2016). ღრმა დასწავლის უახლესი მიღწევები ერთარხიან საველე მასალებში მრავალხმიანი ჩანაწერის თითოეული ხმის ბგერათა სიხშირეების იდენტიფიცირების საშუალებას იძლევა (Bittner et al. 2017; Cuesta, McFee, and Gómez 2020), რაც მასშტაბურ კორპუსებში და ისტორიულ კოლექციებში ჰარმონიული იტერვალების გამოყენების შესწავლის ახალ შესაძლებლობებს გვაძლევს.

წინამდებარე კვლევა ნაწილია პროექტისა „ქართული ტრადიციული მუსიკის ანალიზისა და მოდელირების კომპიუტერული სისტემა საარქივო საველე ჩანაწერებზე დაყრდნობით“ (SRNSFG FR-21-2066), რომელიც მიზნად ისახავს შეიმუშავოს სისტემები მრავალი-F0 დადგენისა და მრავალხმიანი ვოკალური კორპუსების მუსიკალური ინფორმაციის ამოღების (MIR) მეთოდებზე დაფუძნებული ანალიზისთვის.

ჩვენი კვლევა სამშრიანია. პირველ რიგში, ჩვენ ვამუშავებთ თბილისის სახელმწიფო კონსერვატორიის ფოლკლორის ლაბორატორიის არქივიდან აღებულ ვრცელ, გეოგრაფიულად მრავალფეროვან ნიმუშთა ჯგუფებს და ვეძებთ წინასწარი დამუშავების პროცედურულ ჩანაწერებს სეგმენტაციისთვის, ფილტრაციისა და სტანდარტიზაციისთვის. შემდეგ ვაკეთებთ მულტიფუნდამენტური სიხშირის (multi-F0) ალგორითმის ადაპტირებას და გაწრთვნას მრავალხმიანი ჩანაწერებისთვის და თითო ჩანაწერზე ვაგენერირებთ ჰარმონიული ინტერვალების განაწილებებს ცენტრებში კონვერტირებისა და ნარეველი მოდელირების საშუალებით (Scherbaum, Müller, and Rosenzweig 2017; Rosenzweig, Scherbaum, and Müller 2019). ბოლოს კი, ამ განაწილებებიდან ვეძებთ მთლიანი კორპუსის გამოსახულებას, სადაც მონაცემები გამოსახულია დაბალი განზომილების სივრცეში (PCA და UMAP მეთოდებით), ვადგენთ სტრუქტურას HDBSCAN-ის¹ გამოყენებით, ხოლო გამოყვანილი დბალგანზომილებიანი გამოსახულებების მუსიკოლოგიურ მართებულობას ვამოწმებთ ამ გამოსახულებებით რეგი-

1 HDBSCAN — იერარქიულ სიმჭიდროვეზე დაფუძნებული კლასტერიზაციის ალგორითმი, რომლის მეშვეობითაც ხდება სხვადასხვა სიმკვრივის სტაბილური კლასტერების გამოვლენა.

ონული კუთვნილების და ჩანაწერის ფირის ნომრის პროგნოზირების ეფექტურობის დადგენით.

კვლევის კონტექსტი და წინაპირობები

ეთნომუსიკოლოგიური კვლევები ასახავს უნიკალურ ვერტიკალურ აზროვნებასა და ინტერვალურ შედგენილობას საქართველოს სხვადასხვა რეგიონში. განსაკუთრებით გამოარჩევენ კვარტების, კვინტების, ნეიტრალური ტერციებისა და ოქტავეების როლს და ისეთ კულტურულად სპეციფიკურ თანხმომავნებას, როგორიცაა ე.წ. „ეროვნული ტრიქორდი“ 1-4-5 (არაყიშვილი 1950; Nadel 1933; ჭოხონელიძე 1988; ასლანიშვილი 1954). გამოთვლითი კვლევების ადრეულ ეტაპზე, ტონალური ორგანიზაციის კვლევის მიზნით, გამოიყენებოდა მაღალი ხარისხის სტუდიური ჩანაწერები ან ინდივიდუალური ხმების ჩანაწერები (Scherbaum, Müller, and Rosenzweig 2017; Scherbaum et al. 2021, 2022). სვანური სიმღერების შესახებ ჩვენმა წინა კვლევამ აჩვენა, რომ საარქივო აუდიომასალებიდან მრავალი-F0 გამოთვლა ხელს უწყობს ინტერვალური განაწილების ანალიზსა და ჩანაწერების კლასტერიზაციას (Shugliashvili et al. 2024). აღნიშნული კვლევა განზოგადებულია წინამდებარე ნაშრომში, რამდენადაც იგი ეფუძნება ეროვნული მასშტაბის ჩანაწერების უფრო ვრცელ კორპუსს და აერთიანებს თანამედროვე რეპრეზენტაციების შესწავლის და არაზედამხედველობითი შესწავლის მეთოდებს.

თბილისის სახელმწიფო კონსერვატორიის ფოლკლორის ლაბორატორიის არქივი შეიცავს საექსპედიციო მასლების 200-ზე მეტ მაგნიტოფირს, რომლებიც ჩაწერილია 1948-1973 წლებში. მთლიანობაში ქრონომეტრაჟი აღემატება 130 საათს, რომლის დაახლოებით სამი მეოთხედი ვოკალური მუსიკაა. ფირები თავდაპირველად გაციფრებული იყო ერთ WAV ფაილებად, რომლის ცალკეული სექციები შესაბამისი მარკერებით იყო მონიშნული. ჩვენ ავტომატურად ვაქციეთ ეს ფრაგმენტები ინდივიდუალურ ჩანაწერებად დავასათაურეთ (მაგ., „57-9-alilo.wav“) და დავურთეთ შესაბამისი ხელმისაწვდომი მეტამონაცემები (რეგიონი, თემი, ჯგუფი, თარიღი). საკვლევი ბაზიდან გამოვრიცხეთ ინსტრუმენტული ნაწარმოებები, მონოფონიური ჩანაწერები, არაქართული რეპერტუარი, ფრაგმენტული მასალა და ჩანაწერები რეგიონული კუთვნილების გარეშე. კვლევისთვის შერჩეული დარჩენილი მასალა მოიცავს 2,107 მრავალხმიან ვოკალურ ჩანაწერს 18 რეგიონიდან. ეს მასშტაბი ინარჩუნებს რეგიონულ მრავალფეროვნებას და კორპუსის დონის სტატისტიკის შეგროვების საშუალებას იძლევა.

მეთოდები მოდელის განხილვა

ჩვენ ვიყენებთ მიმანიშნებლობაზე დაყრდნობილ არქიტექტურას, რომელიც ეფუძნება DeepSalience და Polyvocals სისტემებს (Bittner et al. 2017; Cuesta, McFee, Gómez 2020), და ვამუშავებთ harmonic constant-Q გარდაქმნას (HCQT) ფაზა-დიფერენციალურ მახასიათებლებთან ერთად, რათა მივიღოთ დრო-სიხშირის მიმანიშნებლობის რუკა $S[t, f]$, რომელიც წარმოაჩენს t დროს ალბათობას, რომ სიხშირე f გაუღერებული ტონის ძირითადი სიხშირეა.

წინამდებარე ნაშრომზე იტერირებით, სტანდარტული VGG-სტილის CNN ბლოკებს ვანაცვლებთ ერთმანეთზე დაწყობილი U-Net-ის ტიპის მოდულებით, რათა გავაუმჯობესოთ მრავალმასშტაბიანი კონტექსტის ერთიანი აღქმა. დამხმარე კავშირები (residual connections) უზრუნველყოფს მოდულის გაწვრთნის სტაბილურობას; გვიანი შერწყმა კი აერთიანებს ამპლიტუდისა და ფაზის შტოებს საბოლოო პროგნოზირების თავის მიღებამდე.

მრავალი-F0 ამომცნობი მოდელის გაწვრთნას ესაჭიროება დიდი რაოდენობის ინდივიდუალურ-არხებიანი აუდიო მასალა. ინდივიდუალურ-არხებიანი მრავალხმიანი ჩანაწერების სიმჭირის გამო, ჩვენ ვქმნით ხელოვნურად მრავალხმიან ნიმუშებს, მონოფონური ჩანაწერების სიმულირებულ ოთახში მოქცევის Pyroomacoustics-ის საშუალებით (Scheibler, Bezzam, and Dokmanić 2018). ჩანაწერთა ჯამების ერთბაშად რევერბერაციის ნაცვლად, თითოეული ხმის ვირტუალურ ოთახში განთავსება უფრო მეტად ემსგავსება ერთმიკროფონიან საექსპედიციო ჩანაწერებს. ღია მონაცემთა ნაკრებების სხვადასხვა მონოფონური ჩანაწერიდან შექმნილი ფსევდომრავალხმიანი ჩანაწერების გარდა, წვრთნა ასევე იყენებს Polyvocal-ში აქამდე გამოყენებული მრავალხმიანი მონაცემების ღია ნაკრებთა კომბინაციასა და GVM კორპუსის ქართული მასალებიდან აღებული შეზღუდული რაოდენობის ნიმუშებს სისტემის ქართულ ჩანაწერებზე ადაპტაციის მიზნით (Cuesta, McFee, and Gómez 2020; Scherbaum 2016).

წვრთნის შემდეგ, მოდელს ვიყენებთ ჩვენს სამიზნე კორპუსში არსებული ჩანაწერებიდან მრავალი-F0 ინფორმაციის ამოსაღებად. ყოველი ჩანაწერიდან ვიღებთ მიმანიშნებლობის რუკას, რომელიც ყოველი დროის მონაკვეთში აჩვენებს ჩანაწერში არსებული ხმების ფუნდამენტურ სიხშირეებს. ამის შემდეგ ვრიცხავთ ანომალიურ უკიდურესობებს (სიხშირეებს $<100\text{ჰც}$ და $>3000\text{ჰც}$) და ამასთან ერთად ვიყენებთ ამბიტუსის გამათვალისწინებელ ფილტრს, რომელიც რიცხავს ჩანაწერში ხმების საშუალო სიხშირეზე ერთ ოქტავაზე მეტით დაშო-

რებულ სიხშირეებს. ამის შემდეგ, F0-ებს გავდარქმნით ცენტებად $A1 = 55$ ჰც-ს მიმართ და ვირჩევთ მხოლოდ სტაბილურ სეგმენტებს სიხშირის ტრაექტორიის სტაბილურობის ნილბის მეშვეობით (Rosenzweig, Scherbaum, and Müller 2019).

თითოეული ჩანაწერის ყოველ დროის მონაკვეთში, გამოვითვლით ყოველ თანმიმდევრულ ვერტიკალურ ინტერვალს ნაპოვნი F0-ებიდან (მაგ.: დროის მონაკვეთში რომელშიც გვხვდება F0-ები [1400 ცენტი, 1650 ცენტი, 1750 ცენტი] A1-თან მიმართებით, ვითვლით ინტერვალებს [250 ცენტი, 100 ცენტი]); ჩანაწერში ამ ხერხით ამოღებულ ვერტიკალურ ინტერვალებს ვათავსებთ ამ ჩანაწერის ვერტიკალური ინტერვალების ნაკრებში. შერბაუმის მეთოდოლოგიის შესაბამისად (Scherbaum, Müller, and Rosenzweig 2017), ნაკრების ალბათობის სიმკვრივის ფუნქციის მოდელირებას ვაკეთებთ ბირთვული სიმჭიდროვის შეფასების მეთოდის მეშვეობით (KDE - kernel density estimation) და დადგენილი ფუნქციიდან ვიღებთ გამორჩეული მწვერვალების რაოდენობას. ამის შემდეგ, ნაკრებს ვარგებთ გაუსის ნარევის მოდელს (GMM - Guassian Mixture Model), რომელშიც კომპონენტების რაოდენობას ვუთანაბრებთ მწვერვალების რაოდენობას; კომპონენტების საშუალო მნიშვნელობები ქმნის ჩანაწერის მნიშვნელოვანი ინტერვალების შემაჯამებელ ნუსხას, ხოლო ნარევის სიმკვრივის ფუნქციის მეშვეობით ვიღებთ უწყვეტი ალბათობის მოდელს. საბოლოოდ, გამოთვლილი ნარევის სიმკვრივიდან 200 თანაბრად დაშორებული ცენტური პოზიციიდან ვიღებთ მნიშვნელობებს, ფიქსირებული სიგრძის ვექტორის მისაღებად, რომელიც გამოიყენება შემდგომი ანალიზისათვის (Scherbaum, Müller, and Rosenzweig 2017; Shugliashvili et al. 2024).

თავდაპირველად ვაკეთებთ ძირითადი კომპონენტების ანალიზს (PCA), რათა დავადგინოთ მონაცემების წრფივი სტრუქტურა და დისპერსიის ახსნადობა. შედეგად: მონაცემების პირველი ორი კომპონენტი აღწერს ცვლადობის 19%-ს, ხოლო პირველი 25-90%-ს (სხვა სიტყვებით, მცირე რაოდენობის კომპონენტებიც კი ასახავს მონაცემების ძირითად განსხვავებებს). არაწრფივი გეომეტრიის გამოსავლენად ვიყენებთ UMAP²-ს 2 და 25 განზომილებაში (McInnes et al. 2018).

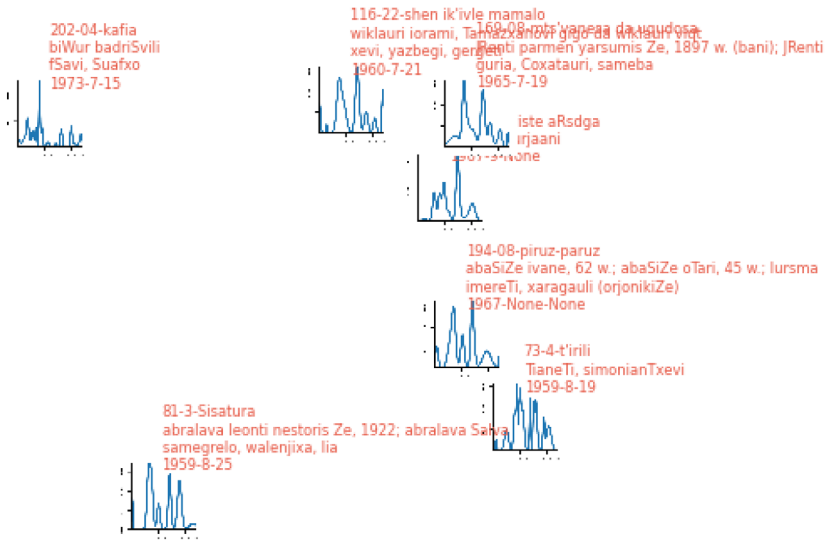
2 UMAP — ერთგვარი ალგორითმი მონაცემთა შეკუმშული (დაბალგანზომილებიანი) გამოსახულების შესაქმნელად, რომელშიც შენარჩუნებულია მათი სტრუქტურა.

PCA კომპონენტების რაოდენობა	დისპერსიის ახსნილი პროპორცია
2	19%
3	27%
5	40%
10	62%
20	85%
25	90%

ცხრილი 1: ჰარმონიული ინტერვალების განაწილებებს (200-ელემენტიან GMM ვექტორებს) შორის დისპერსიის ახსნადობა PCA-ს კომპონენტების მიერ

ჩვენ სტრუქტურა დავადგინეთ HDBSCAN-ის გამოყენებით 2-D UMAP ჩაშენებებზე, რის შედეგადაც მივიღეთ 105 კლასტერი, რომელთაგან 648 ელემენტი მარკირებულია, როგორც ხმაური (დაუკლასტერებელი) (McInnes, Healy, and Astels 2017).

თვისობრივი შესწავლისთვის, ჩვენ ვქმნით კლასტერული მასშტაბების სინოპტიკურ გამოსახულებებს; ამისათვის, თავდაპირველად გამოგვყავს გაუსის ნარევის მოდელი (GMM) თითოეული კლასტერის გაერთიანებული ინტერვალების ნაკრებისთვის. მაგალითად, ერთი კლასტერი, რომელშიც გურული სიმღერების ჩანაწერები ჭარბობს, ავლენს პიკებს 188, 297, 353, 464 და 694 ცენტის ფარგლებში, ხოლო სხვა კლასტერში, რომელშიც ჭარბობს კახური სიმღერების ჩანაწერები, ვლინდება პიკები 192, 298, 352, 414, 571, 658 და 725 ცენტის ფარგლებში, საიდანაც ჩანს, რომ ამ კლასტერებს ინტერვალთა ტერცია საერთო აქვს, თუმცა განსხვავდება უფრო დიდი ინტერვალებით.



სურათი 1: მცირე მონაკვეთი ჩაშენების გრაფიკიდან, სადაც ნაჩვენებია ფერით გამოყოფილი მცირე კლასტერი (კლასტერი N17) ჩანაწერებისა, რომელთა ჰარმონიული ინტერვალების განაწილებები მსგავსია.

რეგიონული ინფორმაციის გამოვლენის მიზნით, გამოწმენდით, ასახავენ თუ არა ინტერვალური განაწილებები კონკრეტული რეგიონების თავისებურებებს. ამისათვის, ჩვენ გამოვიყენეთ ორი გზა. პირველი — მოდალური ეტიკეტების გადატანა: თითოეულ ჩანაწერს ენიჭება მის HDBSCAN კლასტერში ყველაზე ხშირი რეგიონის ეტიკეტი; ამ მეთოდით დაახლოებით 33%-იანი სიზუსტით ხდება ჩანაწერის ეტიკეტირება 18 რეგიონის ფარგლებში. მეორე — კონტროლირებადი კლასიფიკაცია: 80/20 დაყოფით (მონაცემთა შემთხვევითობის პრინციპით არჩეული 80%-ის წვრთნისთვის, დარჩენილი 20%-ის ტესტირებისთვის) ვწვრთნით 100 ხისგან შემდგარ შემთხვევით ტყეებს (random forest) და ვცდილობთ რეგიონის პროგნოზირებას PCA2D, UMAP2D, PCA25D და UMAP25D ჩაშენებების საფუძველზე, რათა დავადგინოთ, თუ რამდენად ინარჩუნებს ეს ჩანარები ამა თუ იმ ჩანაწერში რეგიონულ თავისებურებებზე ინფორმაციას. ამ დავალებაში, UMAP25D აღწევს 40%-იან სანდოობას, თანაბარი სიზუსტით და სენსიტიურობით დაახლოებით 0.4-თან.

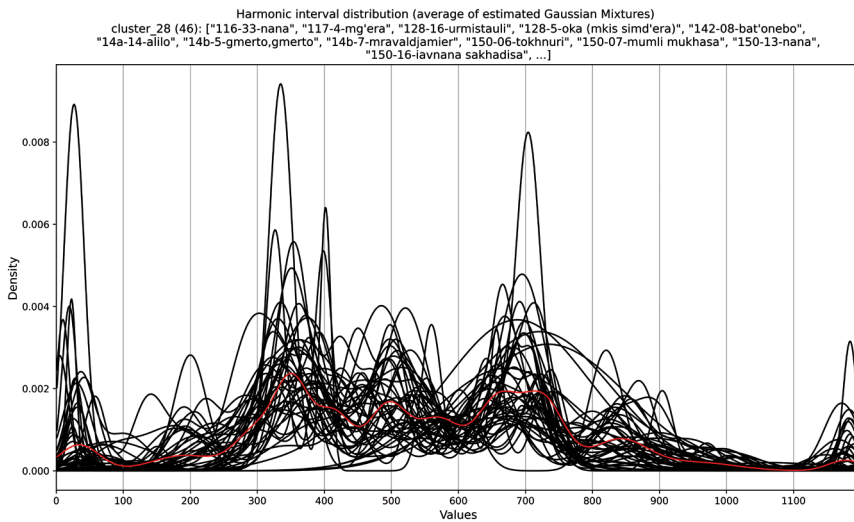
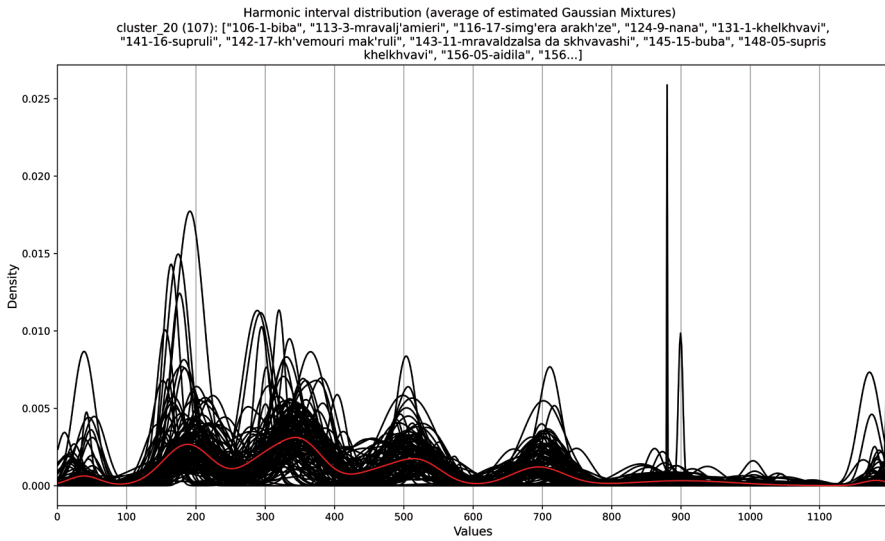
შედეგები

UMAP2D გამოსახულებებზე HDBSCAN-ის მეშვეობით გამოიყო 105 კლასტერი; 648 ჩანაწერი ვერცერთ კლასტერში ვერ მოხვდა. თითოეული კლასტერისთვის განისაზღვრა: ა) ჩანაწერების მოდალური რეგიონი და ბ) GMM-ის კომპონენტების საშუალო მნიშვნელობები. ეს მნიშვნელობები მივიღეთ კლასტერის ჩანაწერების ჰარმონიული ინტერვალების განაწილებიდან ხელახალი ნიმუშების შექმნისა და GMM-ის თავიდან მორგების გზით. შედეგად მივიღეთ სინოპტიკური გამოსახულება, რომელიც აჩვენებს, რა ტიპის ინტერვალები გვხვდება ამ კლასტერის ჩანაწერებში.

კლასტერული ანალიზის შედეგებზე ვიზუალური დაკვირვება ცხადყოფს, რომ ჰარმონიული ინტერვალების მსგავსი სტრუქტურების მქონე ჩანაწერები მართლაც ერთ კლასტერებში არიან განთავსებული.

მაგალითად, კლასტერი 20 (107 ჩანაწერი; მოდალური რეგიონი: გურია; მოდალური ფირის ინდექსი: 70) გამოირჩევა დაახლოებით 188-, 297-, 353-, 464- და 694-ცენტრიანი ინტერვალებით. ამის საპირისპიროდ, კლასტერი 28 (46 ჩანაწერი; მოდალური რეგიონი: კახეთი; მოდალური ფირის ინდექსი: 151) შეიცავს 192-, 298-, 352-, 414-, 571-, 658- და 725-ცენტრიან ინტერვალებს. მცირე ინტერვალები ემთხვევა კლასტერი 20-ისას, მაგრამ განსხვავებები თვალსაჩინოა სეკუნდაზე დიდი ინტერვალების შემთხვევაში.

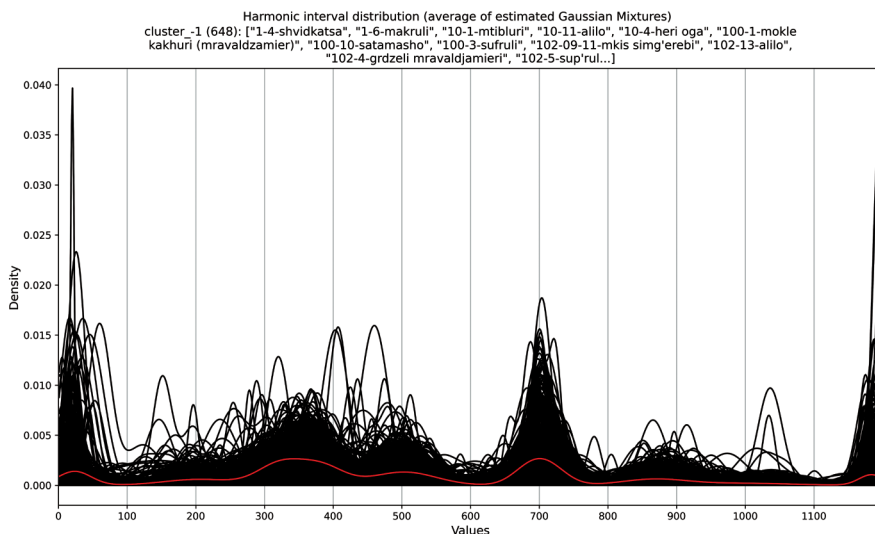
დეტალური მონაცემები – ინტერვალების სინოპტიკური მასშტაბები, მოდალური რეგიონები და ფირების ინდექსები, აგრეთვე სრული ჩაშენების გრაფიკი – ხელმისაწვდომია ჩვენს ბაზაში: https://github.com/lshug/pypolyphonicanalysis_data/tree/main/KADMOS2025.



სურათი 2-3: 28-ე და მე-20 კლასტერში შემავალი ჩანაწერების ჰარმონიული ინტერვალების განაწილებების ერთმანეთზე დალაგებული გრაფიკები. წითელი ხაზები ასახავს ხელახლა მორგებულ GMM-ებს.

„-1“ კლასტერში მოექცა 648 ჩანაწერი, რომლებიც HDBSCAN-მა ვერ მიაკუთვნა კონკრეტულ ჯგუფს. მათი ჰარმონიული ინტერვალე-

ბის განაწილება აჩვენებს ზოგად ტენდენციებს სხვადასხვა რეგიონისთვის საერთო ინტერვალების კუთხით, რაც მიუთითებს კორპუსის ჩანაწერების ერთგვარ საერთო სტრუქტურაზე, თუმცა ეს სტრუქტურა გაცილებით ნაკლებ სისტემურია, ვიდრე დადგენილი კლასტერები.



სურათი 4: HDBSCAN-ის მიერ დაუჯგუფებელი ჩანაწერების ჰარმონიული ინტერვალების განაწილებების ერთმანეთზე დალაგებული გამოსახულება, სადაც ვლინდება კორპუსის მასშტაბით ჰარმონიული ინტერვალების განაწილების საერთო ტენდენციები

რეგიონული მონაცემების წინასწარი პროგნოზირების სიზუსტის შესაფასებლად, ჩვენ დავტესტეთ კლასტერული ჯგუფების პროგნოზირებისა და ჩაშენებების შესაძლებლობები. თავდაპირველად, შევისწავლეთ რამდენად ხშირად ემთხვევა ჩანაწერის კლასტერში ყველაზე ხშირი რეგიონი ჩანაწერის ნამდვილ რეგიონს 18 რეგიონიდან. შედეგად მივიღეთ 33%-იანი სიზუსტე, რაც ნიშნავს, რომ ჩანაწერების მესამედისათვის საკმარისია მხოლოდ იმის ცოდნა, რომელ კლასტერში მოათავსა HDBSCAN-მა მოცემული ჩანაწერი, რათა სწორად გამოვიცნოთ მისი წარმომავლობის რეგიონი.

ამის შემდეგ შევაფასეთ, რამდენად აქვთ თავად ჩაშენებებს ჩანაწერის წარმომავლობის რეგიონის პროგნოზირების უნარი. მონაცემების 80%-20%-იანი ტრენინგ-ტესტ დაყოფის საშუალებით, 100-ხიანი შემთხვევითი ტყეების (random forest) კლასიფიკატორები გაიწვრთნა

ჩანაწერის გეოგრაფიული რეგიონისა და ფირის ინდექსის განსაზღვრის მიზნით UMAP2D, PCA2D, UMAP25D და PCA25D გამოსახულებების საფუძველზე. სიზუსტეები მოცემულია ქვემოთ წარმოდგენილ ცხრილში. განსაკუთრებულად გამოირჩევა ოცდახუთკომპონენტიანი UMAP, რომელმაც მანამდე მისთვი უცნობი ჩანაწერის რეგიონის სწორი კლასიფიცირება შემთხვევების 40%-ში შეძლო. რეგიონის პროგნოზირების მაღალი სიზუსტე ადასტურებს, რომ მიღებული პარმონიული ინტერვალების მახასიათებლები შეიცავს მნიშვნელოვან მუსიკალურ და კულტურულ ინფორმაციას, რაც სანდოობას ანიჭებს ჩვენს გამოთვლით მეთოდს.

ჩაშენების ტიპი	სანდოობა	სიზუსტე	სენსიტიურობა
PCA2D	0.17	0.1	0.1
PCA25D	0.39	0.19	0.2
UMAP2D	0.25	0.15	0.14
UMAP25D	0.4	0.41	0.4

ცხრილი 2: ჩანაწერების რეგიონული კუთვნილების კლასიფიკაციის შედეგები, მიღებული 2-განზომილებიან და 25-განზომილებიან PCA და UMAP ჩაშენებებზე დაყრდნობით

განხილვა

კლასტერების სინოპტიკური სქემების რაოდენობრივი ანალიზი ამყარებს ადრინდელ სამეცნიერო თვალსაზრისს ქართული მრავალზომიანი სიმღერის ვერტიკალური კოორდინაციის შესახებ (Nadel 1933; არაყიშვილი 1950; ჭოხონელიძე 1988; ასლანიშვილი 1954).

105 HDBSCAN კლასტერში, რომლებიც UMAP2D ჩაშენებებიდან მივიღეთ, 103 კლასტერში მინიმუმ ერთი გამოკვეთილი ინტერვალის გვხვდება ტერციის ფარგლებში (300-400 ცენტი). 100 კლასტერში გვხვდება ინტერვალის, რომელიც უახლოვდება წმინდა კვინტას (700 ცენტი); 76 კლასტერში გვხვდება ინტერვალის წმინდა კვარტის (500 ცენტი) ფარგლებში და 93 კლასტერში – ოქტავა (1200 ცენტი). ეს მაჩვენებლები მიუთითებს, რომ ტერციები და კვინტები ვერტიკალური ტექსტურის საყრდენს წარმოადგენს, ხოლო კვარტები და ოქტავები რამდენადმე ნაკლებად უნივერსალურია.

იმ კლასტერის სინოპტიკური სქემა, რომლის წევრების რაოდენობა ყველაზე დიდია ($n = 107$, მოდალური რეგიონი: გურია), ავლენს

პიკებს, რომელთა ცენტობრივი მაჩვენებლებია: 188, 297, 353, 464, 524, 694, 903 და 1181 ცენტები. იგი აერთიანებს ძალზე მჭიდრო და ინტენსიურ ტერციულ სტრუქტურულ რეგიონს კვინტებისა და ოქტავების კომპონენტებით. ეს ტენდენციები შეესაბამება კარგად დოკუმენტირებულ აღწერილობას პრეფერენციული ვერტიკალური ჰარმონიული ურთიერთობების შესახებ (Mzhavanadze and Scherbaum 2020; Scherbaum and Müller 2023; Scherbaum et al. 2022).

მოდალური რეგიონების კლასტერებში განაწილების ფაქტორებს შორის, გარდა რეპერტუარისა და ჩაწერის პირობებისა, მნიშვნელოვანი ადგილი სწორედ რეგიონულ კუთვნილებას უკავია. 105 კლასტერში კახეთი წარმოადგენს 31 კლასტერს (432 ჩანაწერი), გურია — 14 კლასტერს (240 ჩანაწერი), სვანეთი — 11 კლასტერს (139 ჩანაწერი) და სამეგრელო — 11 კლასტერს (138 ჩანაწერი).

კვინტები კახეთის თითქმის ყველა და გურიის აბსოლუტურად ყველა კლასტერში გვხვდება, ხოლო ტერციების პიკები (300-350 ცენტის ფარგლებში) გავრცელებულია ყველა რეგიონში. დიდ სეკუნდაზე ოდნავ პატარა ინტერვალი (188 ცენტი), რომელიც განსაკუთრებით თვალსაჩინოა გურიის ყველაზე დიდ კლასტერში, გვხვდება გურიის კლასტერთა დაახლოებით ნახევარში, თუმცა კახეთში გაცილებით იშვიათია.

UMAP2D-ისა და HDBSCAN-ის გამოყენებით მიღებული კლასტერები ქმნის ასიმეტრიული ტიპის განაწილებას, სადაც ბევრი პატარა, თუმცა თანმიმდევრული ჯგუფია (კლასტერის საშუალო ზომა 10; 75-ე პროცენტული 16; მაქსიმუმი 107) და საკმაოდ დიდი ხმაურის ნაკრები („-1“ კლასტერი, 648 ერთეული). მაღალი კლასტერული დაფარვა მიუთითებს, რომ ინტერვალების განაწილების ვექტორები საიმედოდ ქმნის მჭიდრო სამეზობლოებს, თუმცა საკმაოდ მოზრდილი ხმაურის ფრაქცია ალბათ მრავალფეროვანი ფაქტორების ნაზავს ასახავს: ექსპედიციის სხვადასხვა კონტექსტში აკუსტიკური ცვალებადობა, რეგიონული ნიშნით აღბეჭდილი რეპერტუარის ჰეტეროგენურობა და დაბალგანზომილებიან სივრცეში სიმკვრივეზე დაფუძნებული კლასტერირების კონსერვატიზმი. თითოეულ კლასტერზე საშუალოდ 9 სინოპტიკური პიკის არსებობა (დიაპაზონი 4-14) მიუთითებს, რომ ჩვენი KDE-ის პიკების დათვლისა და GMM-ზე მორგების პროცედურით შესაძლებელი ხდება მნიშვნელოვანი ვერტიკალური ინტერვალების შეზღუდული, მაგრამ ინტერპრეტირებადი კომპლექტების დაჭერა; მიუხედავად ამისა, ერთმანეთთან ახლოს მდებარე ინტერვალები, მათი სიმჭიდროვიდან გამომდინარე, შესაძლოა, ზედმეტად ან არასაკმარისად დაიყოს. მთლიანობაში, ეს ანალიზი ადასტურებს ჩვენს

ადრინდელ თვალსაზრისს: ინტერვალების სტატისტიკა თავისთავად შეიცავს მუსიკალურად მნიშვნელოვან სტრუქტურას, რომელიც შეესაბამება რეგიონულ მეტამონაცემებს, მაგრამ ამავე დროს შესაძლოა, მგრძნობიარე იყოს ჩაწერის პირობებისა და რეპერტუარზე გავლენის მექანე ფაქტორების მიმართ, რაც ცალსახად უნდა იქნას მოდელირებული მომავალ კვლევებში.

დასკვნები და სამომავლო ამოცანები

მრავალზმანიან საარქივო ჩანაწერებიდან პირდაპირ გამოთვლილი ჰარმონიული ინტერვალების განაწილებითი გამოსახულებები მუსიკალურად მნიშვნელოვან ინფორმაციას შეიცავს. მიმანიშნებლობაზე დაფუძნებული მრავალი-F0 გამოთვლა, GMM ინტერვალური მოდელები და თანამედროვე მანიფოლდური სწავლების კომბინაცია ავლენს სტრუქტურას, რომელიც შეესაბამება რეგიონულ მეტამონაცემებს და უზრუნველყოფს პროგნოზირების შესაძლებლობებს 18 რეგიონში საკმაოდ მაღალი სიზუსტით.

მომავალი მიმართულებები მოიცავს: 1) თითოეულ ჩანაწერში ინტერვალების გაურკვევლობის კვანტიფიკაციას; 2) ვერტიკალური და ჰორიზონტალური (მელოდიური) სტრუქტურის ერთობლივ მოდელირებას; 3) სხვადასხვა კორპუსის შედარებასა და დროში ბგერათწყობის პარამეტრების ცვლილების ანალიზს; 4) უფრო ვრცელ და ღრმა მუსიკოლოგიურ ინტერპრეტაციას რეგიონული ექსპერტების ჩართულობით და სვანეთზე ფოკუსირებული კვლევის გაშლა-გაფართოებას (Shugliashvili et al. 2024; Scherbaum, Müller, and Rosenzweig 2017; Scherbaum et al. 2022; Scherbaum and Müller 2023).

მადლობა

წინამდებარე კვლევა განხორციელდა შოთა რუსთაველის საქართველოს ეროვნული სამეცნიერო ფონდის მხარდაჭერით (SRNSFG FR-21-2066). მადლობას ვუხდით ჩვენს პარტნიორებსა და თბილისის სახელმწიფო კონსერვატორიის ფოლკლორის ლაბორატორიას მათი ფასდაუდებელი წვლილისათვის.

დამოწმებანი

- არაყიშვილი, დიმიტრი. 1950. *სვანური ხალხური სიმღერები*. თბილისი: გამომცემლობა „ხელოვნება“.
- ასლანიშვილი, შალვა. 1954. „მრავალხმიანობის ფორმები ქართულ ხალხურ სიმღერებში“. *კრებულში ნარკვევები ქართული ხალხური სიმღერების შესახებ* ტ.1, 3-89. თბილისი: გამომცემლობა „ხელოვნება“.
- ახობაძე, ვლადიმერ. 1957. *ქართული (სვანური) ხალხური სიმღერები*. თბილისი: შრომა და ტექნიკა.
- ვეშაპიძე, ლევან და ზაალ წერეთელი. 2014 „ქართული ბგერათაწყობის შესახებ“. *ტრადიციული მრავალხმიანობის მეშვიდე საერთაშორისო სიმპოზიუმი*. სექტემბერი 22-26, თბილისი.
- ერქვანიძე, მალხაზ. 1992. „ქართული მუსიკის წყობა“. *ხელოვნება* 7-8: 66-80.
- ჭობონელიძე, ევსევი. 1988. „ხმათა კოორდინაციისა და ჰარმონიის ზოგიერთი თავისებურება ქართულ ხალხურ მრავალხმიან (პოლიფონიურ) მუსიკაში“. *კრებულში ქართული მუსიკის პოლიფონიის საკითხები, რედ.-ები ნოდარ გაბუნია, რუსუდან წურწუმია და სხვანი*, 29-42. თბილისი: თბილისის სახელმწიფო კონსერვატორია.
- Аракишвили, Дмитрий. 1905. *Краткий очерк развития грузинской карталино-кахетинской народной песни с приложением нотных примеров и 27 песен в народной гармонизации*. Москва: тип. К.Л. Меньшова.
- Bittner, Rachel M., Brian McFee, Justin Salamon, Peter Li, and Juan P. Bello. 2017. “Deep Salience Representations for F0 Estimation in Polyphonic Music.” In *Proceedings of the International Society for Music Information Retrieval Conference*.
- Cuesta, Helena, Brian McFee, and Emilia Gómez. 2020. “Multiple F0 Estimation in Vocal Ensembles Using Convolutional Neural Networks.” In *Proceedings of the 21st International Society for Music Information Retrieval Conference*, 302-309. Montréal.
- McInnes, Leland, John Healy, Nathaniel Saul, and Lukas Großberger. 2018. “UMAP: Uniform Manifold Approximation and Projection.” *Journal of Open Source Software* 3 (29): 861.
- McInnes, Leland, John Healy, and Steve Astels. 2017. “hdbscan: Hierarchical Density Based Clustering.” *Journal of Open Source Software* 2 (11): 205.
- Mauch, Matthias, and Simon Dixon. 2014. “pYIN: A Fundamental Frequency Estimator Using Probabilistic Threshold Distributions.” In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 659-663. Florence, Italy.

- Mzhavanadze, Nana, and Frank Scherbaum. 2020. "Svan Funeral Dirges (Zär): Musicological Analysis." *Musicologist* 4 (2): 168-197.
- Nadel, Siegfried F. 1933. *Georgische Gesänge*. Leipzig: Harrassowitz.
- Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. 2015. "U-Net: Convolutional Networks for Biomedical Image Segmentation." arXiv:1505.04597.
- Rosenzweig, Sebastian, Frank Scherbaum, and Meinard Müller. 2019. "Detecting Stable Regions in Frequency Trajectories for Tonal Analysis of Traditional Georgian Vocal Music." In *Proceedings of the International Society for Music Information Retrieval Conference*.
- Scheibler, Robin, Eric Bezzam, and Ivan Dokmanić. 2018. "Pyroomacoustics: A Python Package for Audio Room Simulations and Array Processing Algorithms." In *Proceedings of IEEE ICASSP*. Calgary.
- Scherbaum, Frank. 2016. "On the Benefit of Larynx-Microphone Field Recordings for the Documentation and Analysis of Polyphonic Vocal Music." In *Proceedings of the 6th International Workshop on Folk Music Analysis*. Dublin.
- Scherbaum, Frank, Nana Mzhavanadze, Sebastian Rosenzweig, and Meinard Müller. 2022. "Tuning Systems of Traditional Georgian Singing Determined from a New Corpus of Field Recordings." *Musicologist* 6 (2): 142-168.
- Scherbaum, Frank, and Meinard Müller. 2023. "From Intonation Adjustments to Synchronization of Heart Rate Variability: Singer Interaction in Traditional Georgian Vocal Music." *Musicologist* 7 (2): 155-177.
- Scherbaum, Frank, Meinard Müller, and Sebastian Rosenzweig. 2017. "Analysis of the Tbilisi State Conservatory Recordings of Artem Erkomaishvili in 1966." In *Proceedings of the 7th International Workshop on Folk Music Analysis*. Málaga.
- Shugliashvili, Levan, Nana Mzhavanadze, David Shugliashvili, and Frank Scherbaum. 2024. "Tracking Changes in Harmonic Interval Structures of Svan Songs Based on Computational Analysis of Field Recordings from 1959-1971." *History, Archaeology, Ethnology* XI: 654-670.
- Чхиквадзе, Г. З. 1964. *Основные типы грузинского народного многоголосья*. Москва: Наука.

Levan Shugliashvili

Independent researcher, Georgia
levanspublicemail@gmail.com

David Shugliashvili

Tbilisi State Conservatoire, Georgia
david.shugliashvili@tsc.edu.ge

Nana Mzhavanadze

Tbilisi State Conservatoire, Georgia
nana.mzavanadze@tsc.edu.ge

Frank Scherbaum

Geophysics at Institute of Geosciences
University of Potsdam, Germany
fs@geo.uni-potsdam.de

DOI: 10.32859/kadmos/17/7-32

Harmonic Interval Organization in Georgian Traditional Polyphony: Evidence from Computational Analysis of Archival Field Recordings (1948–1973)*

Keywords: computational ethnomusicology, multi-F0 estimation, harmonic intervals, Georgian polyphony, archival audio

Introduction

Georgian polyphonic singing features complex voice coordination and a vertical orientation that scholars have described for over a century (Nadel 1933; Akhobadze 1957; Chokhnelidze 2010/1988; Chkhikvadze 2010/1964; Aslanishvili 2010/1954). Theoretical debates persist around the nature of tuning and intervallic organization in regional traditions, including Svane-ti, Guria, Kakheti, and others (Аракишвили 1905; Arakishvili 2010/1950;

* This research was supported by the Shota Rustaveli National Science Foundation of Georgia (SRNSFG FR-21-2066).

Veshapidze and Tsereteli, 2014; Mzhavanadze and Scherbaum 2020; Scherbaum and Müller 2023; Scherbaum et al. 2022). Until recently, corpus-scale acoustical analysis of vertical intervals depended on data with per-voice F0 annotations (e.g., the Erkomaishvili recordings; Scherbaum, Müller, and Rosenzweig 2017) or specialized capture modalities such as larynx microphones (Scherbaum 2016). Advances in deep learning now enable estimating multi-voice pitch salience directly from single-channel field recordings (Bittner et al. 2017; Cuesta, McFee, and Gómez 2020), opening new pathways to study harmonic interval usage at scale and across historical collections.

This study is part of the project “Computational System for Analysis and Modelling of Georgian Traditional Music Based On Archival Field Recordings” (SRNSFG FR-21-2066), aimed at developing systems for multi-F0 estimation and music information retrieval (MIR)-based analysis of polyphonic vocal corpora. Our contribution is threefold. First, we process a large, geographically diverse subset from the TSC Folklore Laboratory archive and design a preprocessing pipeline to segment, filter, and standardize recordings. Second, we adapt and train a multiple fundamental frequency estimator for polyphonic singing and derive per-recording harmonic interval distributions via cent-domain processing and mixture modeling (Scherbaum, Müller, and Rosenzweig 2017; Rosenzweig, Scherbaum, and Müller 2019). Third, we build corpus-level representations from these distributions, learn low-dimensional embeddings (PCA, UMAP), discover structure via HDBSCAN, and probe musicological validity through region prediction tasks and label transfer.

Background

Ethnomusicological accounts document distinctive vertical thinking and intervallic inventories across Georgian regions, highlighting the roles of fourths, fifths, neutral thirds, and octaves, as well as culturally specific sonorities such as the so-called 1–4–5 “national trichord” (Arakishvili 2010/1950; see also discussions in Nadel 1933; Chokhonelidze 2010/1988; Aslanishvili 2010/1954). Computationally, early work leveraged high-quality studio recordings or per-voice captures to study tonal organization (Scherbaum, Müller, and Rosenzweig 2017; Scherbaum et al. 2021; 2022). Our previous study focusing on Svan songs demonstrated that multi-F0 estimation from archival audio can support distributional analysis of intervals and clustering of recordings (Shugliashvili et al. 2024). The present work generalizes these ideas to a broader national corpus and integrates modern representation learning and unsupervised discovery.

The TSC Folklore Laboratory archive comprises more than 200 digitized expedition tapes recorded between 1948 and 1973, totaling over 130 hours of audio, roughly three-quarters of which are vocal music. Tapes were initially digitized as single WAV files with cue markers. We automatically segmented them into individual items, normalized filenames (e.g., “57-9-alilo.wav”), and associated available metadata (region, community, group, date). We excluded instrumentals, monophonic items, non-Georgian repertoire, fragments, and items lacking regional attribution. The resulting analysis subset contains 2,107 polyphonic vocal recordings from 18 regions. This scale enables corpus-level statistics while preserving regional diversity.

Methods Model overview

We adopt a salience-based architecture inspired by DeepSalience and Polyvocals, processing a harmonic constant-Q transform (HCQT) along with phase-differential features into a time–frequency salience map $S[t, f]$ representing the probability of a sung pitch at frequency f and time t (Bittner et al. 2017; Cuesta, McFee, and Gómez 2020). Building on prior iterations, we replace standard VGG-style CNN blocks with stacked U-Net-like modules to improve multi-scale context aggregation; residual connections stabilize training; and late fusion combines magnitude and phase branches before the final prediction head.

Training the model requires a larger quantity of multi-channel recordings as the training dataset. To mitigate the scarcity of multitrack choral stems, we synthesize pseudo-polyphonic mixtures by convolving monophonic stems with room impulse responses simulated via pyroomacoustics (Scheibler, Bezam, and Dokmanić 2018). Compared with naive reverberation of sums, per-source placement in a virtual room better approximates single-microphone expedition recordings. In addition to pseudo-polyphonic recordings created from various monophonic recordings from open datasets, training uses a mixture of open polyphonic datasets previously employed in Polyvocals and a limited subset from Georgian materials from the GVM corpus for domain adaptation (Cuesta, McFee, and Gómez 2020; Scherbaum 2016).

After training, we use the model to extract multi-F0 information of the recordings in our corpus. For each recording, we obtain a salience map containing, at each point in time, fundamental frequencies of the voices within the recording at that time. We prune implausible extremes (<100 Hz or >3000 Hz) and apply an ambitus-aware mask removing components more than one octave away from the per-recording mean. We then transform F0s to cents with

respect to $A1 = 55$ Hz, and retain only stable segments using a stability mask (Rosenzweig, Scherbaum, and Müller 2019).

At each timestep of each recording, we compute all successive vertical intervals among detected F0s (e.g., for timestep with F0s at [1400 cents, 1650 cents, 1750 cents] above $A1$, we compute intervals [250 cents, 100 cents]), pooling across time to form a per-recording sample of encountered intervals. Following Scherbaum’s methodology (Scherbaum, Müller, and Rosenzweig 2017), we estimate the probability density function of the sample with Kernel Density Estimation and detect the number of prominent peaks. We then fit a Gaussian mixture model (GMM) with that peak count; the component means form a synoptic list of salient intervals for the recording, while the mixture density provides a continuous probability model. Finally, we sample the estimated mixture density at 200 equally spaced cent positions to obtain a fixed-length vector for downstream analysis (Scherbaum, Müller, and Rosenzweig 2017; Shugliashvili et al. 2024).

We first examine PCA to assess linear structure and variance compression: the first two components explain 19% of the variance and the first 25 explain 90%. To capture non-linear geometry we apply UMAP in two and twenty-five dimensions (McInnes et al. 2018).

Number of PCA components	Explained variance proportion
2	19%
3	27%
5	40%
10	62%
20	85%
25	90%

Table 1: explained variance across harmonic interval distributions (200-element GMM vectors) per number of PCA components

We discover structure using HDBSCAN over the 2-D UMAP embedding, obtaining 105 clusters, with 648 items labeled as noise (unclustered) (McInnes, Healy, and Astels 2017). For qualitative inspection we compute cluster-level synoptic scales by refitting a GMM to the pooled intervals of each cluster. As examples, one Guria-dominant cluster exhibits peaks near 188, 297, 353, 464, and 694 cents, while a Kakheti-dominant cluster features peaks near 192, 298, 352, 414, 571, 658, and 725 cents, echoing shared thirds but diverging in larger intervals.

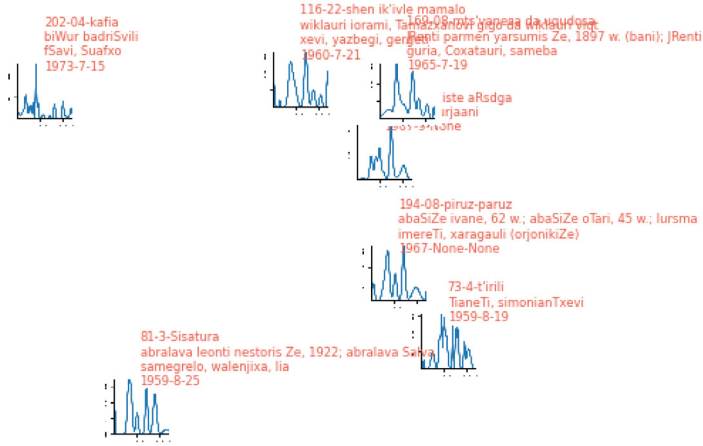


Figure 1: a small section of the embedding plot, showcasing a small color-coded cluster (Cluster 17) of recordings with similar harmonic interval distributions.

To quantify whether interval distributions encode regional information, we run two probes. First, modal-label transfer: each recording inherits the modal region of its HDBSCAN cluster; this yields 33% accuracy over 18 regions. Second, supervised classification: random forests (100 trees) trained with an 80/20 split predict region (and, separately, tape index) from PCA2D, UMAP2D, PCA25D, and UMAP25D embeddings. In this task, UMAP25D achieves 40% accuracy with balanced precision/recall around 0.4.

Results

Clustering UMAP2D representations with HDBSCAN resulted in 105 clusters, with 648 recordings not being placed into any cluster. For each cluster, we identify the modal region of the recordings within the cluster, as well the means of the components of the GMM obtained through re-sampling from the harmonic interval distributions of recordings within the sample and refitting a GMM, representing a synoptic view of the harmonic intervals appearing within the songs in the cluster.

Visual inspection of the results of clustering reveals that distinct groups of recordings with similar harmonic interval structures are indeed placed within the same clusters. For example, Cluster 20 (107 recordings, modal region: Guria, modal tape index: 70) exhibits characteristic intervals at approximately 188, 297, 353, 464, and 694 cents; in contrast, Cluster 28 (46 recordings, modal region: Kakheti, modal tape index: 151) features intervals at 192, 298, 352, 414,

571, 658, and 725 cents, with smaller intervals matching those of Cluster 20 but with deviations in intervals above the second. Detailed cluster data, including the synoptic interval scales, modal regions and tape indices, as well as the full embedding plot, are available in our repository (https://github.com/lshug/pypolyphonicanalysis_data/tree/main/KADMOS2025).

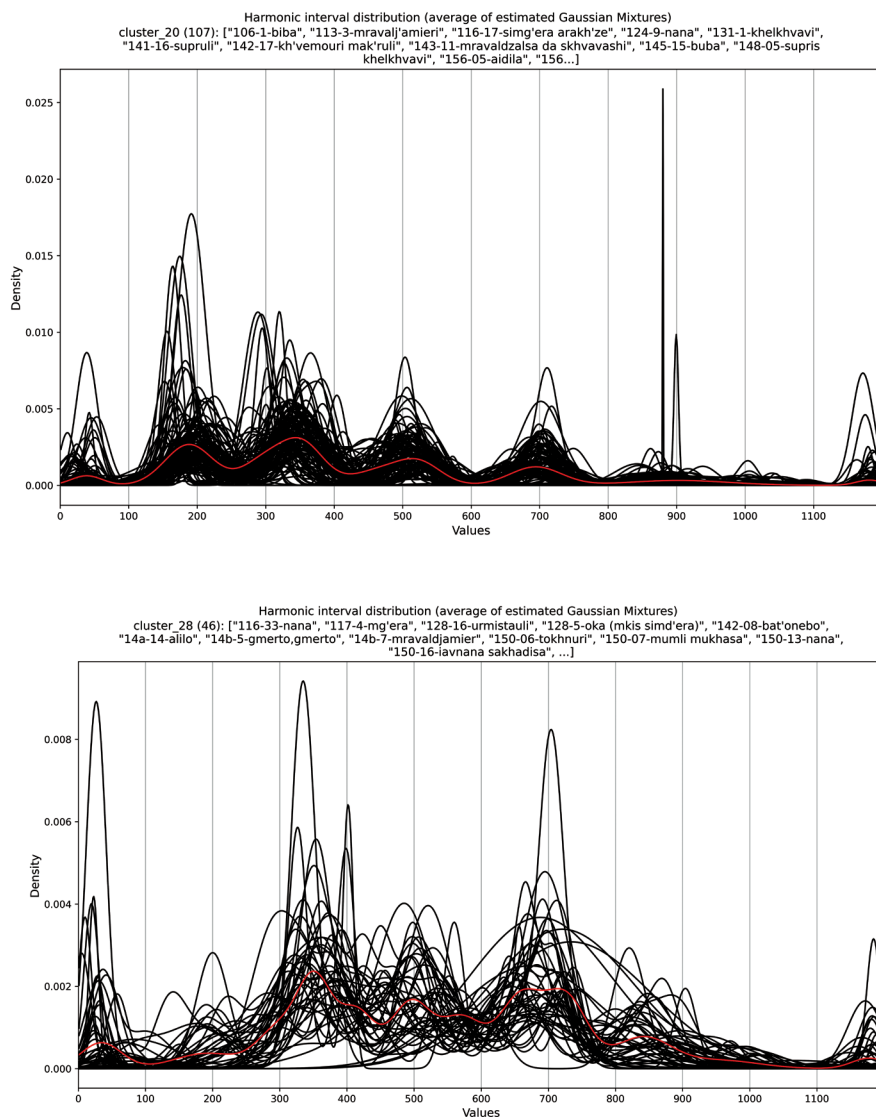


Figure 2-3: Superimposed harmonic interval distributions of recordings in clusters 28 and 20. The red lines represent the refitted cluster GMM.

The “-1” cluster includes 648 recordings unclustered by HDBSCAN. Their harmonic interval distributions show global trends in common intervals, suggesting shared structure across the corpus, though less coherent than identified clusters.

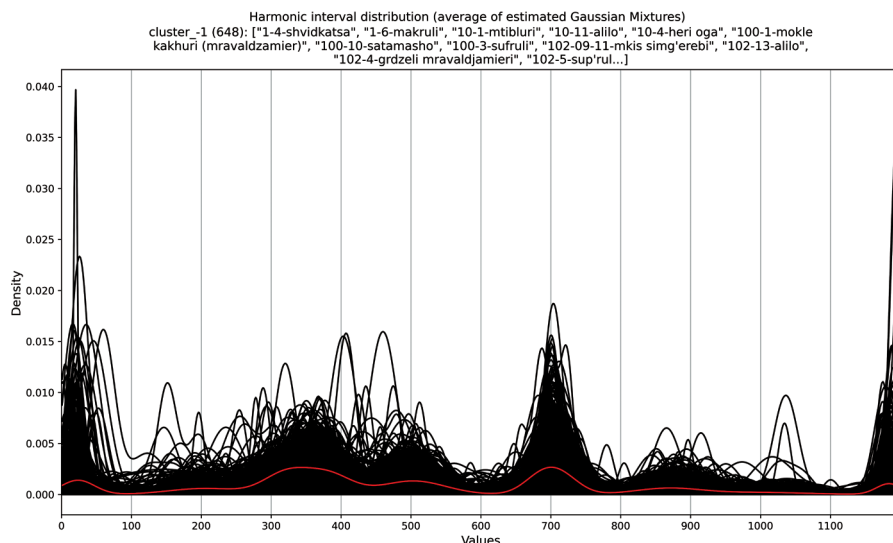


Figure 4: Superimposition of the harmonic interval distributions of the recordings left unclustered by HDBSCAN, revealing corpus-wide patterns in harmonic interval distributions

To assess the robustness of our findings, we test the predictive power of cluster assignments and embeddings. First, we assessed how often the modal region of a recording’s cluster matches the recording’s true region of origin across 18 regions, obtaining the accuracy of 33%, meaning that for a third of the recordings, simply knowing in which cluster HDBSCAN placed the recording is enough to correctly predict the recording’s region of origin.

Following this, we evaluated to what extent the embeddings themselves held predictive power over the region of origin. Using an 80%-20% train-test split, random forest classifiers (100 estimators) are trained to predict geographical regions and tape indices from UMAP2D, PCA2D, UMAP25D, and PCA25D representations of the recordings. Accuracies are presented in the table below, with a standout of 25-component UMAP being able to correctly classify previously unseen recording’s region of origin in 40% of the cases. High classification accuracy for regional predictions confirms that the extracted harmonic interval features encode meaningful musical and cultural information, validating our computational approach.

Embedding	Accuracy	Precision	Recall
PCA2D	0.17	0.1	0.1
PCA25D	0.39	0.19	0.2
UMAP2D	0.25	0.15	0.14
UMAP25D	0.4	0.41	0.4

Table 2: Results of recording region classification based on 2-dimensional and 25-dimensional PCA and UMAP embeddings

Discussion

Quantitative summaries from the cluster-level synoptic scales reinforce prior accounts of vertical coordination in Georgian multipart singing (Nadel 1933; Arakishvili 2010/1950; Chokhnelidze 2010/1988; Aslanishvili 2010/1954). Across the 105 HDBSCAN clusters obtained from the UMAP2D embedding, 103 clusters include at least one prominent mode in the “third region” (~300–400 cents), 100 include a mode near the perfect fifth (~700 cents), 76 include a mode near the perfect fourth (~500 cents), and 93 include an octave (~1200 cents). These rates indicate that thirds and fifths are ubiquitous supports in the vertical texture, with fourths and octaves also common but somewhat less universal. The cluster with the most members ($n = 107$, modal region Guria) exhibits a synoptic scale with peaks near ~188, ~297, ~353, ~464, ~524, ~694, ~903, and ~1181 cents, combining dense third-region structure with strong fifth and octave components—patterns consistent with well-documented descriptions of preferred vertical relations (Mzhavanadze and Scherbaum 2020; Scherbaum and Müller 2023; Scherbaum et al. 2022).

The distribution of modal regions over clusters suggests meaningful regional organization alongside repertoire and session effects. Kakheti accounts for 31 of 105 clusters (432 recordings), Guria for 14 (240), Svaneti for 11 (139), and Samegrelo for 11 (138). Fifths appear in nearly every Kakheti cluster and all Guria clusters, while third-region peaks are widespread across regions. A lower-second feature near 188 cents, conspicuous in the largest Guria cluster, occurs in roughly half of Guria clusters but is much rarer in Kakheti.

UMAP2D and HDBSCAN yields a long-tailed partition with many small but coherent groups (median cluster size 10; 75th percentile 16; maximum 107) and a substantial noise set (648 items in the “-1” cluster). The high clustering coverage indicates that interval-distribution vectors reliably form dense neighborhoods, but the sizeable noise fraction likely reflects a mixture of fac-

tors: acoustic variability across expedition contexts, repertoire heterogeneity within region labels, and the conservatism of density-based clustering in low dimensions. The average of nine synoptic peaks per cluster (range 4-14) suggests that our KDE to peak-count to GMM procedure captures a limited but interpretable set of salient vertical intervals; nevertheless, closely spaced modes can be under- or over-split depending on local density. Together, these observations support our earlier claims: interval statistics alone encode musically meaningful structure aligned with regional metadata, while potentially remaining sensitive to session- and repertoire-level confounds that future work should model explicitly.

Conclusions and future work

Distributional representations of harmonic intervals, estimated directly from archival polyphonic recordings, encode musically meaningful information. The combination of salience-based multi-F0 estimation, GMM interval models, and modern manifold learning reveals structure aligned with regional metadata and supports predictive probes at non-trivial accuracy across 18 regions. Future directions include (i) per-recording uncertainty quantification for intervals; (ii) joint modeling of vertical and horizontal (melodic) structure; (iii) cross-corpus comparisons and time-series analyses of tuning change; and (iv) richer musicological interpretation in dialogue with regional experts, expanding upon earlier Svan-focused work to a nation-scale synthesis (Shugliashvili et al. 2024; Scherbaum, Müller, and Rosenzweig 2017; Scherbaum and Müller 2023).

Acknowledgments

This work was supported by the Shota Rustaveli National Science Foundation of Georgia (SRNSFG FR-21-2066). We thank our collaborators and the TSC Folklore Laboratory for their invaluable contributions.

References

- Akhobadze, Vladimer. 1957. *Collection of Georgian (Svan) Folk Songs*. Tbilisi: Shroma da Teknika. ახოზაძე, ვლადიმერ. 1957. *ქართული (სვანური) ხალხური სიმღერები*. თბილისი: შრომა და ტექნიკა.
- Аракишвили, Дмитрий. 1905. *Краткий очерк развития грузинской карталино-кахетинской народной песни с приложением нотных примеров и 27 песен в народной гармонизации*. Москва: тип. К.Л. Меньшова.
- Arakishvili, Dimitri. 2010 (originally published in 1950 in Georgian). "Svan Folk Song." In Tsurtsunia and Jordania 2010, 35-57.
- Aslanishvili, Shalva. 2010 (originally published in 1954 in Georgian). "Forms of Multipart Singing in Georgian Folk Songs." In Tsurtsunia and Jordania 2010, 57-81.
- Bittner, Rachel M., Brian McFee, Justin Salamon, Peter Li, and Juan P. Bello. 2017. "Deep Salience Representations for F0 Estimation in Polyphonic Music." In *Proceedings of the International Society for Music Information Retrieval Conference*.
- Chkhikvadze, Grigol. 2010 (originally published in 1964 in Russian). "Main Types of Georgian Folk Polyphony." In Tsurtsunia and Jordania 2010, 97-110.
- Chokhonelidze, Evsevi. 2010 (originally published in 1988 in Georgian). "Some Characteristic Features of Voice Coordination and Harmony in Georgian Multipart Singing." In Tsurtsunia and Jordania 2010, 135-145.
- Cuesta, Helena, Brian McFee, and Emilia Gómez. 2020. "Multiple F0 Estimation in Vocal Ensembles Using Convolutional Neural Networks." In *Proceedings of the 21st International Society for Music Information Retrieval Conference*, 302-309. Montréal.
- Erkvanidze, Malkhaz. 1992. "The Sound Structure of Georgian Music." *Khelovneba* (Art) 1992, nos. 7-8: 66-80. Tbilisi. ერქვანიძე, მალხაზ. 1992. „ქართული მუსიკის წყობა.“ *ხელოვნება* 7-8: 66-80.
- McInnes, Leland, John Healy, Nathaniel Saul, and Lukas Großberger. 2018. "UMAP: Uniform Manifold Approximation and Projection." *Journal of Open Source Software* 3 (29): 861.
- McInnes, Leland, John Healy, and Steve Astels. 2017. "hdbscan: Hierarchical Density Based Clustering." *Journal of Open Source Software* 2 (11): 205.
- Mauch, Matthias, and Simon Dixon. 2014. "pYIN: A Fundamental Frequency Estimator Using Probabilistic Threshold Distributions." In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 659-663. Florence, Italy.

- Mzhavanadze, Nana, and Frank Scherbaum. 2020. "Svan Funeral Dirges (Zär): Musicological Analysis." *Musicologist* 4 (2): 168-197.
- Nadel, Siegfried F. 1933. *Georgische Gesänge*. Leipzig: Harrassowitz.
- Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. 2015. "U-Net: Convolutional Networks for Biomedical Image Segmentation." arXiv:1505.04597.
- Rosenzweig, Sebastian, Frank Scherbaum, and Meinard Müller. 2019. "Detecting Stable Regions in Frequency Trajectories for Tonal Analysis of Traditional Georgian Vocal Music." In *Proceedings of the International Society for Music Information Retrieval Conference*.
- Scheibler, Robin, Eric Bezzam, and Ivan Dokmanić. 2018. "Pyroomacoustics: A Python Package for Audio Room Simulations and Array Processing Algorithms." In *Proceedings of IEEE ICASSP*. Calgary.
- Scherbaum, Frank. 2016. "On the Benefit of Larynx-Microphone Field Recordings for the Documentation and Analysis of Polyphonic Vocal Music." In *Proceedings of the 6th International Workshop on Folk Music Analysis*. Dublin.
- Scherbaum, Frank, Nana Mzhavanadze, Sebastian Rosenzweig, and Meinard Müller. 2022. "Tuning Systems of Traditional Georgian Singing Determined from a New Corpus of Field Recordings." *Musicologist* 6 (2): 142-168.
- Scherbaum, Frank, and Meinard Müller. 2023. "From Intonation Adjustments to Synchronization of Heart Rate Variability: Singer Interaction in Traditional Georgian Vocal Music." *Musicologist* 7 (2): 155-177.
- Scherbaum, Frank, Meinard Müller, and Sebastian Rosenzweig. 2017. "Analysis of the Tbilisi State Conservatory Recordings of Artem Erkomaishvili in 1966." In *Proceedings of the 7th International Workshop on Folk Music Analysis*. Málaga.
- Shugliashvili, Levan, Nana Mzhavanadze, David Shugliashvili, and Frank Scherbaum. 2024. "Tracking Changes in Harmonic Interval Structures of Svan Songs Based on Computational Analysis of Field Recordings from 1959-1971." *History, Archaeology, Ethnology* XI: 654-670.
- Tsurtsumia, Rusudan and Joseph Jordania. 2010. *Echoes from Georgia: Seventeen Arguments on Georgian Polyphony*. New York: Nova Science.
- Veshapidze, Levan, and Zaal Tsereteli. 2014. "About the Georgian traditional sound system." *Seventh International Symposium on Traditional Polyphony*, September 22-26, Tbilisi, Georgia. ვეშაპიძე, ლევან და ზაალ წერეთელი. 2014. „ქართული ბგერათაწყობის შესახებ: ტრადიციული მრავალხმიანობის მეშვიდე საერთაშორისო სიმპოზიუმი. სექტემბერი 22-26, თბილისი.