Supplemental Material: MAICO: A Visualization Design Study on AI-Assisted Music Composition

Simeon Rau, Frank Heyen, Benedikt Brachtel, and Michael Sedlmair *Member, IEEE*

I. Introduction

This supplemental document contains more extensive versions of subsections from the main paper's design section. Here, we provide more information on the similarity and emotional metrics we use and how they influence specific emotions, with which we create layouts. Furthermore, we present our sample representations with additional layout examples: We showcase all glyphs for different types of monophonic samples to demonstrate strengths and drawbacks. Additionally, we include some of the figures as bigger versions to show their details.

II. Design

Given our tasks and goals from our paper, we present more extensive information on our design, as well as additional images, that illustrate the design in different use cases.

A. Filters

Using the filter from Figure 1a), composers could define a pitch range. For all notes where the pitch is over the maximum that is allowed, we move it down by 12 steps (one octave) until it is in the range. For notes under the minimum, we move it up by one octave until the pitch fits the range.

When defining the chromas using the filter Figure 1b), we move all contradicting pitches up by one step until they reach a pitch that is allowed.

All these changes were suggested by the professional composer Brachtel.

B. Correlation

Composers might be interested in the correlation/relationship of two metrics or model parameters or use these to arrange samples in space. This allows for ordered results regarding the chosen metrics to spot outliers or groups of samples that users like to investigate. For example, users might want to find samples with many notes that have a high similarity to the primer or analyze the influence of the temperature parameter on the similarity to the primer melody.

In order to get an overview of the possible combinations and their correlation, we provide a matrix visualization with color-coded cells and show the pairwise Pearson correlation (Figure 2). The cells' color ranges from red (negative correlation) over white (no correlation) to blue (positive correlation), guiding the user toward a more thoughtful selection of axes, as users can predict the impact directly without inspecting all combinations of parameters and metrics. We also provide a single-hue coloring so it is easy to find extreme values (Figure 2 bottom left). By clicking a cell, for example, the number of notes and temperature, the main view shows the selected metrics on the axis and places all samples according to the values. Figure 2 shows one example of a positive correlation between temperature used during generation and the number of notes in resulting samples.

C. Emotion-Based Similarity

We chose features to express emotion as an alternative to our similarity metric. To this feature table we applied dimensionality reduction to map all samples to 2D. To make this choice, we researched existing work [1]–[3], [5]–[11], to evaluate, how our metrics influence emotions. The metrics we chose and their influence are shown in Table I.

D. Similarity-Based Layout

As shown in our paper, we display a large number of samples using our similarity metrics for placement. While paper already shows an image of the melodic similarity based layout, also with gridified version Figure 3 (a, b), we added emotion-based placements (optionally gridified) for the same data in Figure 3 (c, d).

E. Glyphs

We presented our different approaches to visualizing samples in our paper. Here, we now want to provide more information about glyph designs here by comparing them on different characteristic melodies to demonstrate their advantages and weaknesses (Table II, see Figure 4 for a legend).

When only notes of a *single pitch* occur, the piano roll shows a straight line and the chroma roll a single hue. The metric flower can reveal the number of notes (top petal), while the complexity flower and the rhythm pie show that there are no rests.

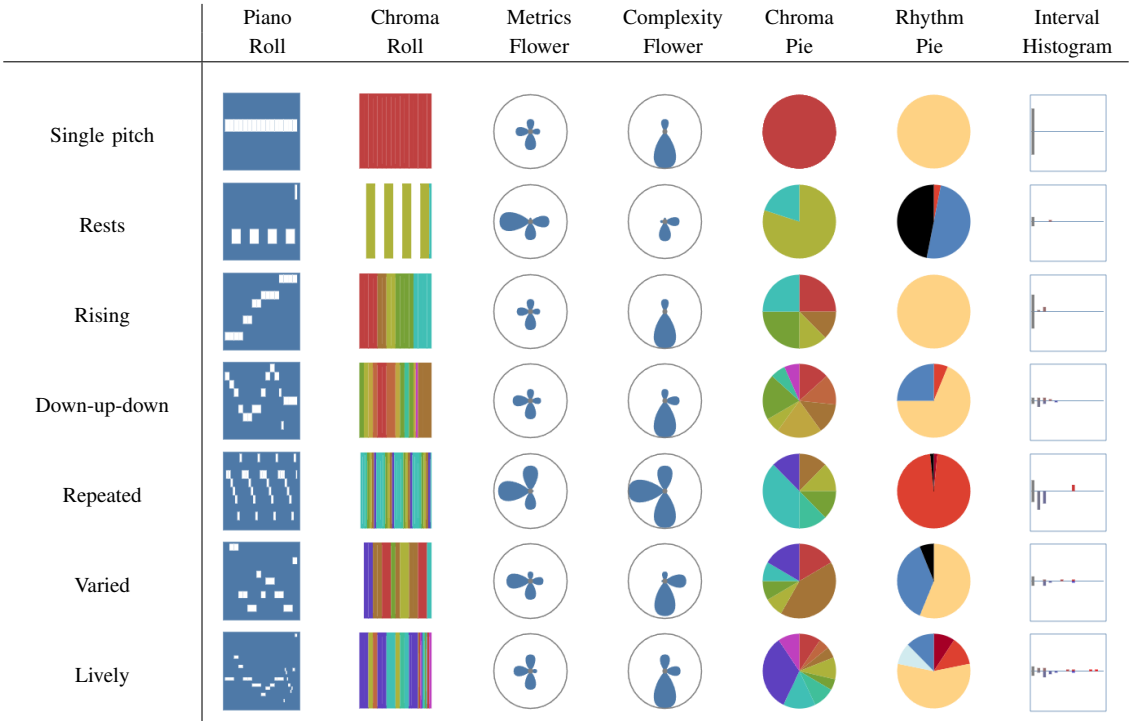
If a melody contains *rests*, the complexity flower's bottom petal is small the rhythm pie contains a big black area. The positions of the rests are shown in both the piano and chroma roll with empty spaces between notes.

Rising melodies show up in piano rolls as a rising streak of blocks and in the chroma roll as a smooth change in hues. The rhythm pie reveals that there are still only quarter notes (same

TABLE I: Influence of metrics on the estimated emotion of the melody: Positive (+), negative (-), and no influence (blank).

Metric	Happy	Sad	Excited	Calm	Suspense
Major key	+	-		+	
Minor key	-	+		-	
Rhythm complexity			+	-	
Number of notes	+	-	+	-	
Range of notes	+	-	+	-	
Mean of intervals	+	-	+	-	
Number of perfect intervals	+	+		+	
Number of non-perfect intervals	-			-	+
Mean pitch	+	-	+	-	
Ascending melody			+		+
Descending melody		+		+	

TABLE II: Comparison of glyph types for different examples of melody samples. All of them are actual outputs of the models we used.



as the previous example), while the jump histograms reveal that there are only minor and major seconds in the rising part.

The melodic structure of *down-up-down* can be easily seen in the piano roll. This melody has some rhythmic changes, which can make the rhythm interesting, which is shown in the complexity flower to the right. Compared to the previous examples, the rhythm pie reveals the different note lengths. We can quickly derive that this melody has some notes that are not in the scale of the key because the chroma pie shows eight slices.

Repeated melodic patterns show up as repeated shapes and hue-sequences in piano and chroma rolls. We can see that this melody also has a small rest, which is indicated by the complexity flower (bottom) and the rhythm pie's black slice. We can also see in the metric flower (right) that the mean length of notes is very short.

Varied melodies show different pitches, intervals, and rhythmic changes. The jump histogram shows directly all occurring intervals.

The *lively* melody combines all the previously mentioned aspects and shows a variety of intervals, note lengths, and pitches while showing reoccurring patterns in the chroma roll.

For example, by comparing all chroma pies, we can derive that all melodies use the same pitches from a limited pool, which in this case are all notes from the C major scale plus one additional pitch. Especially C, D, F, G, and A are often present throughout all melodies where some occur more often in specific ones like A in the *lively* melody.

1) Aggregated Model Summaries: In order to provide an overview of a model (T5) as a whole, we provide small glyphs to represent aggregated information (Figure 5). Similar to the cluster representative glyphs, this overview allows users to get an impression without deep analysis, so they might quickly discard, like, or inspect a model directly.

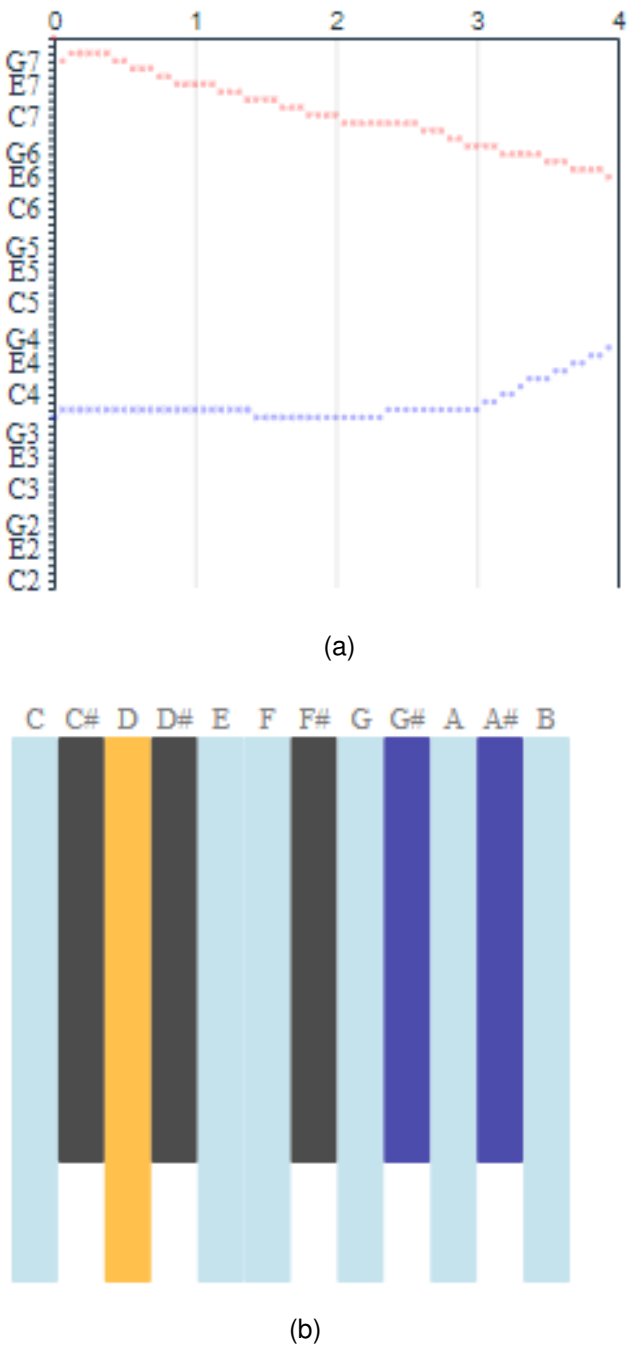


Fig. 1: Two filters allow steering model sampling (a) by range and (b) by chroma. In this example, (a) we narrow the range towards the end to get a controlled and more concentrated ending. Also, (b) we allow any chroma but C♯, D♯, and F♯ (gray), while we select D as our root note (orange) to look for samples that are highly related to the D minor key using our other visualizations, as Brachtel did in his commissioned work.

We chose the chroma pie and the metric flower, as the occurring chromas and the general similarity to the primer or variance of intervals could be interesting to find models that often produce variations or lively samples.

We also show small versions of the correlation matrix for each model, so users can quickly find and compare interesting differences and correlations (Figure 5a).

In order to represent all samples at once, we use our aggregated density piano roll glyph that shows how often a note occurs at each time step and pitch (Figure 5b). This visualization allows users to see the diversity a model produces or whether there is a time step where the same note is played in every snippet.

As mentioned above, we track users' actions to facilitate recalling seen and heard samples and allow users to label the samples. We show these statistics in the model summary, so users can quickly detect unexplored, disliked, and liked models when approaching a decision.

F. Clusters

Due to our method of producing polyphonic samples, the same single voice can occur in many samples, allowing composers to exclude regions if the voice is not interesting. We, therefore, connect every pair of samples that shares the same monophonic sample with a colored line (Figure 6). We chose a set of differentiable colors to indicate the ten biggest clusters, as excluding one of these thins out the space more efficiently than excluding small ones while also allowing us to find a lot of interesting samples faster. To avoid clutter in dense regions, we apply edge bundling [4].

G. Harmonics

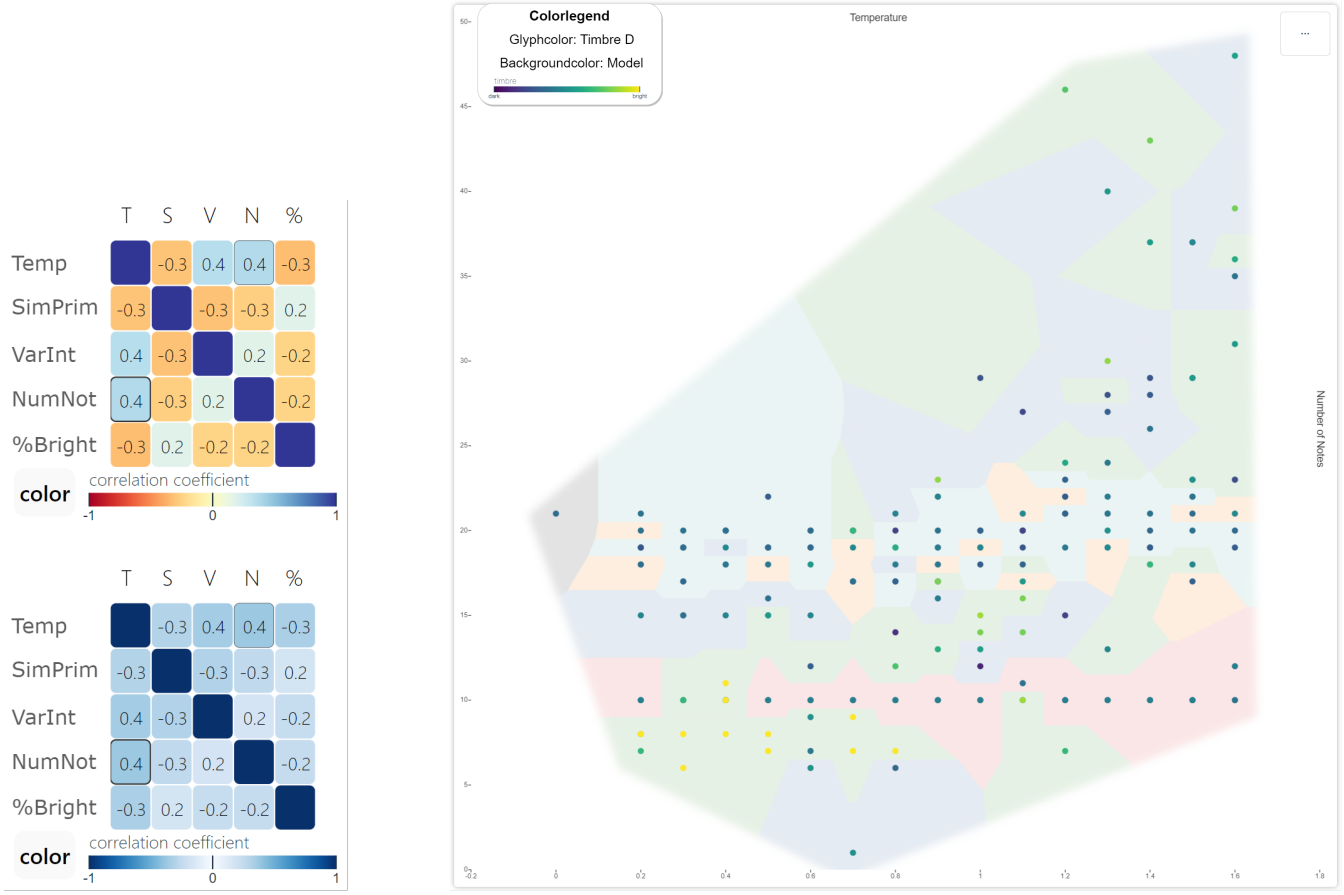
We use a parallel coordinate plot, where the root notes are used as features and the timbre as value to determine the positions. Thus, each sample occurs 12 times (instead of once like in other layouts) and is colored by timbre from dark (purple) to bright (yellow) – encoding timbre brightness as color brightness. This plot allows composers to follow the sentiment of a single snippet over multiple root notes, as well as selecting interesting regions based on timbre, ultimately selecting samples that could fit a sentiment or intention. We position the root notes in the sequence of the circle of fifths, as this represents its calculation best and reveals interesting patterns (Figure 7).

III. Evaluation

As described in section III-A (Design-Process), we evaluated our initial prototype on model comparison. In the following, we present more details on the process and results of this study.

A. Initial Study

To further evaluate how users understand and interact with our design, we conducted a pair analytics [**?**] study with four musicians. In pair analytics, the interface designers assist the participants with help and hints, allowing them to use all



(a) (b)

Fig. 2: The correlation matrix color-codes correlation between pairs of parameters/metrics in two modes: **Top left:** A diverging color scale allows differentiating between positive and negative correlation. **Bottom left:** A single-hue scale for magnitude facilitates comparing the strength of correlation between matrix cells. **Right:** The overview visualization for the axis selected in (a) – temperature and number of notes. There is indeed a positive correlation, as samples that were generated with higher temperatures tend to have more notes.

features correctly and effectively without having to learn them first. Our participants had different levels of musical knowledge: P1 and P3 are hobby musicians with over five years of experience, P2 studies an instrument at a music university and composes occasionally, and P4 is a professional full-time jazz bassist. We met with each participant individually for 75 to 120 minutes, and recorded voice, computer audio, and the screen. The session began with a short introduction of the general concept, models and parameters, and visual interface.

Our participants generally understood the visuals but found it *"overwhelming at the beginning"*1 and hard to understand how metrics, encodings (P3), and visualizations (P4) would translate to musical insights. As MAICO is an expert tool, this was expected and the reason why we did pair analysis, as we could help them understand and answer their questions. Despite the initial learning curve, all participants offered highly positive feedback on the visual approach, as our design choices help to get an overview of large spaces with *"many different melodies [...] visualization is very important to find specific [melodies]"*2 and made them *"like to explore multiple options and analyze what fits best"*1.

P2 wanted to know why all melodies of one model were *"concentrated in the similarity [plot]"*2 and found *"this model uses the exact same notes from the primer"*2 using our glyphs. The density piano rolls showed *"many melodies are the same, and there is a bit of variation"*2. A common strategy of P2 and P3 was investigating very different melodies to previous findings. To this end, P2 used chroma pies and metric layouts to find melodies with different pitches compared to the primer. Another strategy was to use the metric-layout with number of notes and temperature to find models that produce a lot of notes, some of which are not included in the primer. In conclusion, this sequence led to insights about one model, and P2 found two models fitting his personal preference.

Another common strategy was analyzing interesting regions or melodies based on their goals. The filters helped looking for melodies with specific characteristics like *"small intervals [...] and not too many notes"*1. Furthermore, P3 used the combination of metric-based scatterplot and filters to look for melodies similar to the primer but with a few out-of-

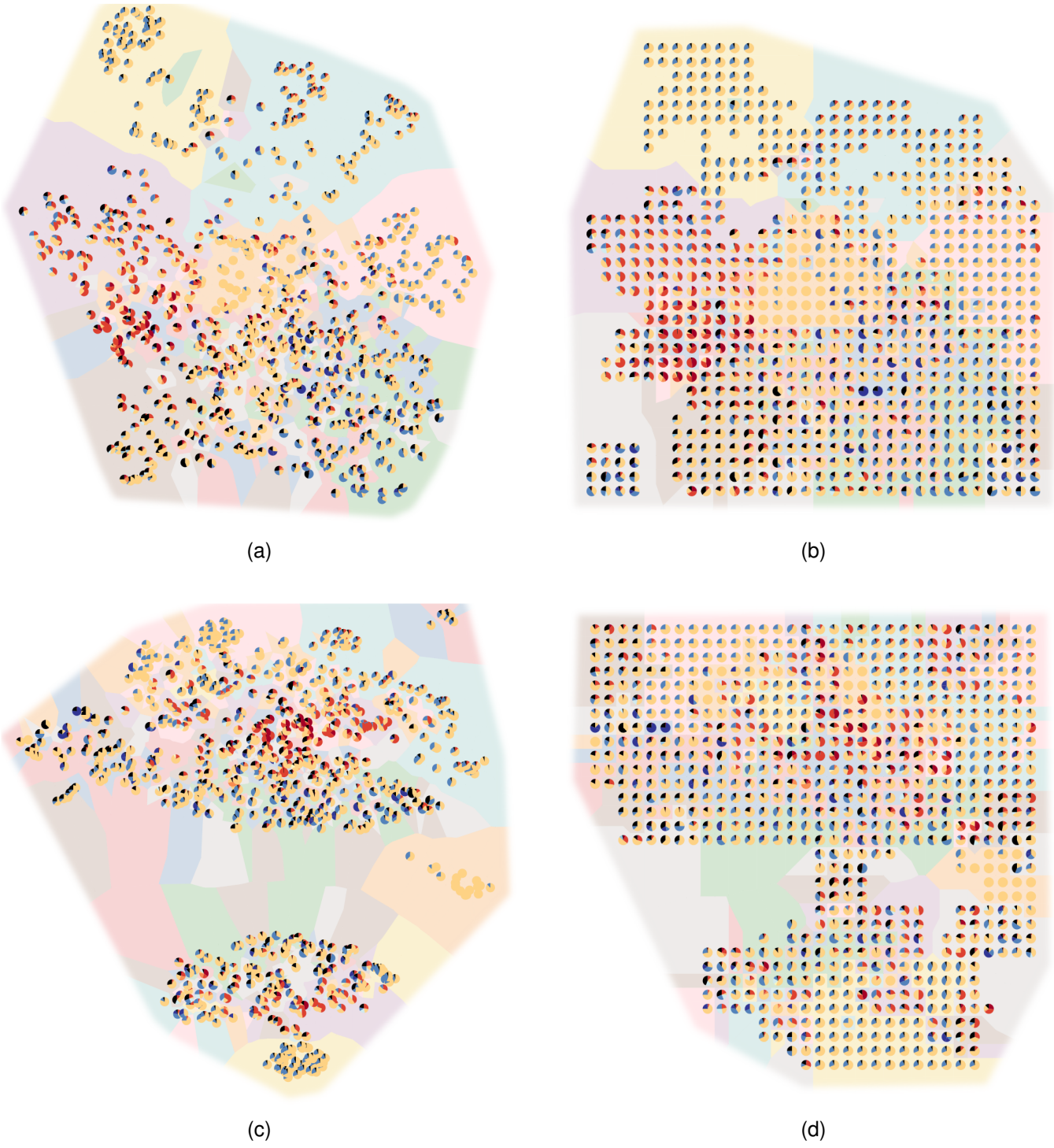


Fig. 3: Four different options for a metric-based layout showing rhythm pie glyphs. The glyphs are positioned by either melodic similarity **(a, b)** or emotional similarity **(c, d)**. The resulting layout can be either exact **(a, c)** or gridified to avoid overlap **(b, d)**. In this example, each glyph's background color encodes the model that generated the sample. In the emotion-based plots (bottom), the mixed background colors indicate that different models are more intertwined and that emotion does not separate the models as clearly as melodic similarity. Instead, we can see two strongly separated emotion clusters in **(c)**, where the bottom one shows mostly melodies from the purple and yellow model. We note that these clusters are harder to separate in the gridified layout **(d)**, which instead fosters better readability of individual sample glyphs.

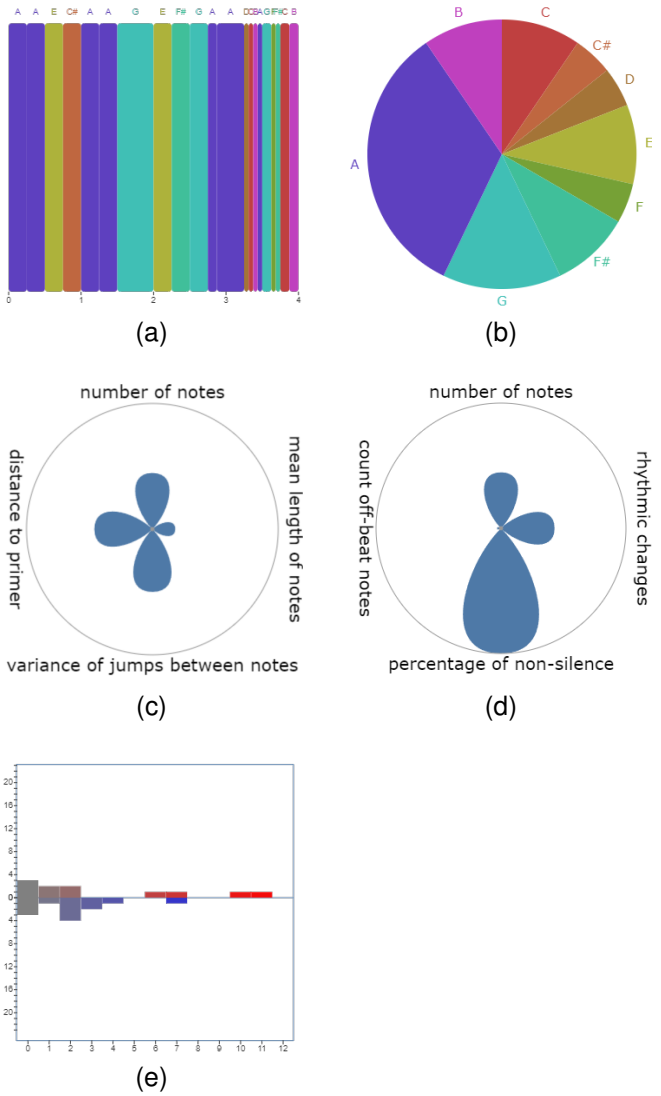


Fig. 4: The different types of glyphs that represent melody samples: **a)** chroma rolls encode pitch as color, **b)** chroma pie charts show chroma/duration distributions, **c)** metric & **d)** rhythmic complexity flower glyphs show different pitch/rhythmic metrics, and **r)** interval histograms show (up/down) pitch jump distributions.

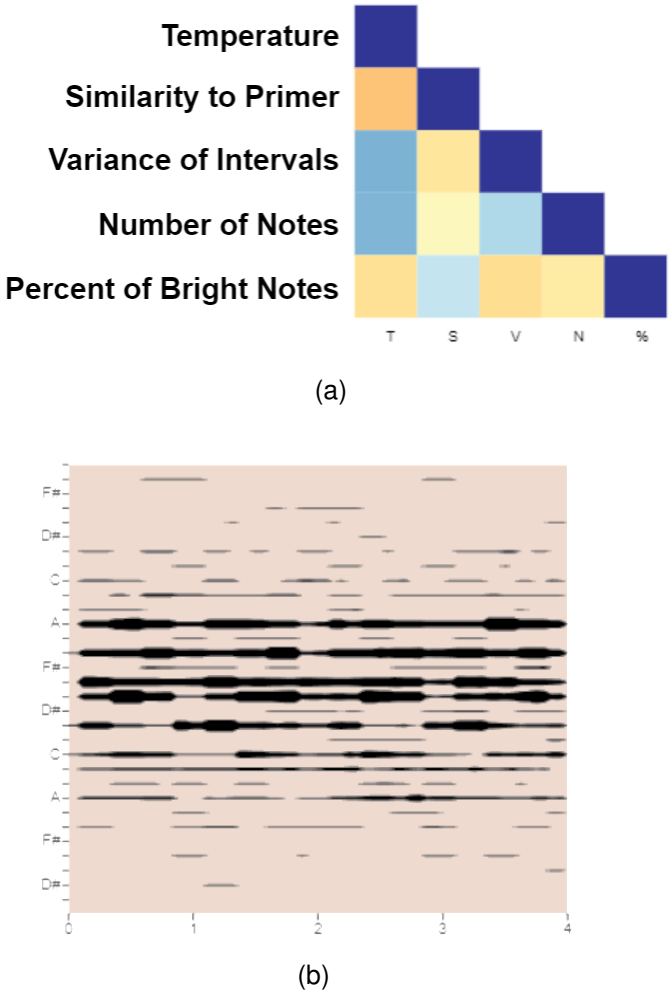


Fig. 5: Our model glyphs show aggregated information from all samples that were generated by that model. We also use flower and chroma pie glyphs for this purpose, which are not shown here since they are similar to the ones for single samples (Figure 4).

the-ordinary accents, as these *"out-of-scale note makes this [melody] interesting"*3. Others investigated regions with single models, but results were*"not good enough"*1. Similar to looking at regions, clustering showed P2 a cluster with two models, the blue and red ones. P2 concluded that *"these models are working similarly"*2 as they were not visually separable in the similarity-based layout.

Changing layouts, P2 compared the blue and red models and saw that *"they are not the same, because one model has more similarity to the primer at low temperature"*2, whereas he *"would not have known where differences are [without using visualization]"*2. Using our glyphs, P1 was looking for models that produce slight variations of the primer while P3 looked *"for AIs that do something totally different"*3. Mainly P3 gained insights about models using our metric layouts, as *"the orange [model] is good with a low number of notes but not with many [notes]"*3 or that the pink model *"is clearly trained on classical piano music"*3.

Our participants wondered about the impact of the temperature parameter (sensitivity analysis) and used the correlation matrix for this analysis: *"The purple model seems to generate more notes than the others, even at higher temperatures"*1. *"The higher the temperature, the weirder the melody gets [...], but the red model still produces something good"*3.

P3 stated that our visualizations help with analyzing models, but coming to a decision is not possible without this labeling: *"which AI is good? actually I can't remember"*3, and that tracking would help to see unexplored AIs: *"I did not listen to this AI"*3.

Generally, participants liked all our metrics and glyphs and found cases where they were useful but wished to have some reference visuals and values from the primer (P4).

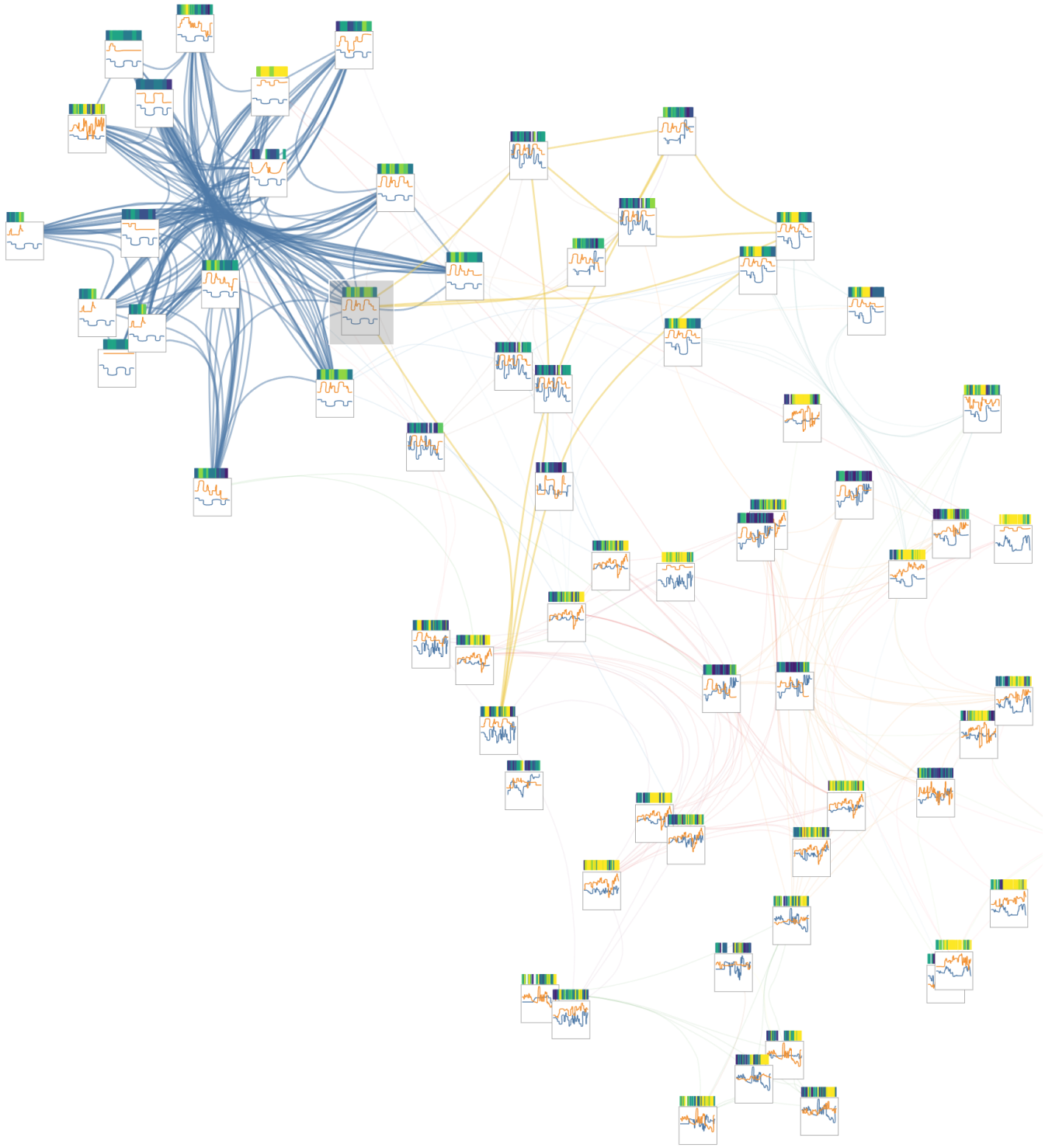


Fig. 6: This representation reveals three big clusters in blue, red, and orange, showing that around 70% of all polyphonic samples share one of these three monophonic samples. Clearly visible is a big cluster of polyphonic samples that share the same voice (blue links), which can all be excluded if their voice is not interesting. The voice lines glyphs show interesting samples with clearly separated voices (top left) and overlapping/crossing voices (bottom right).

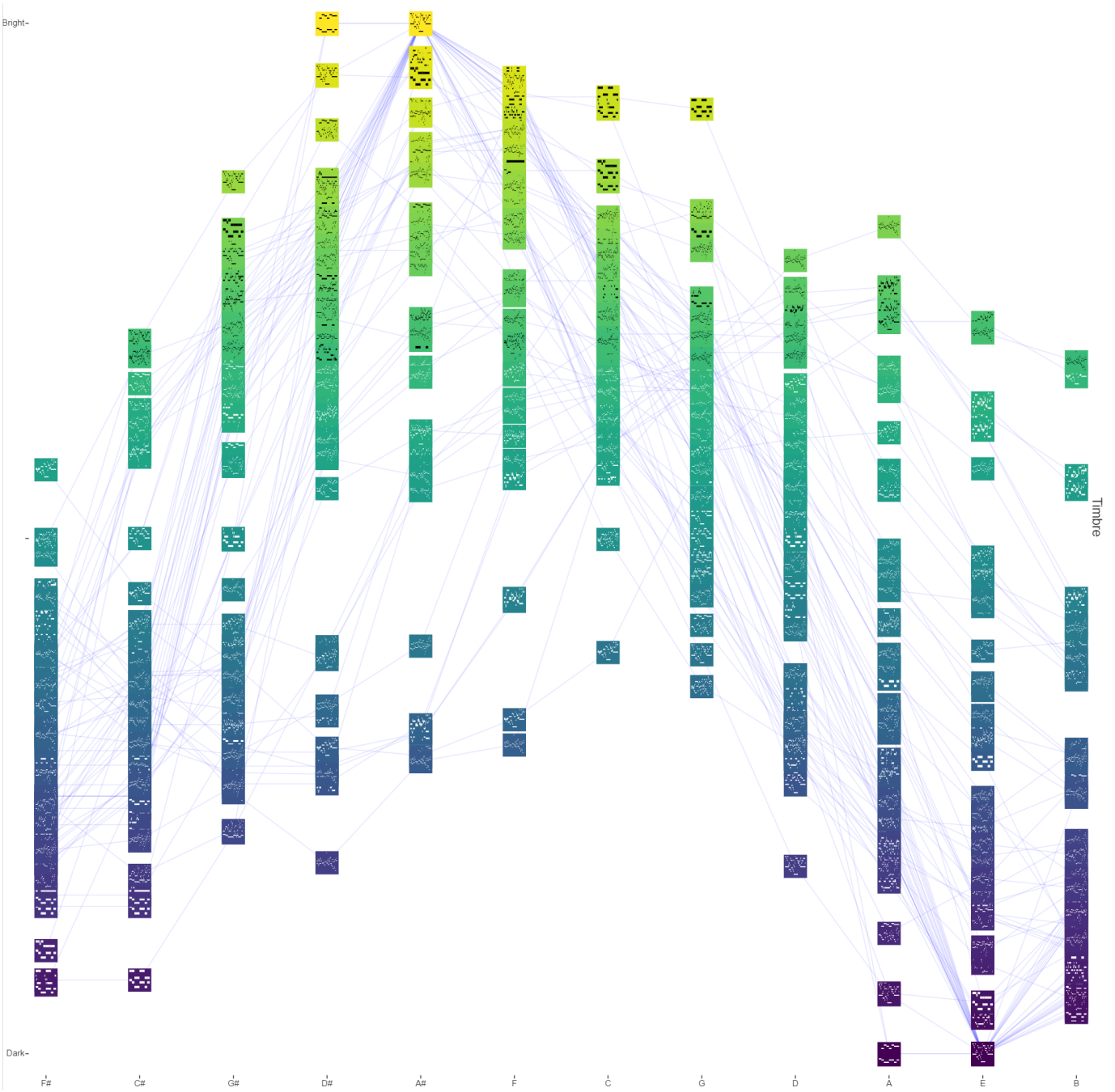


Fig. 7: Harmonic visualization for polyphonic samples using the timbre brightness metric. This example shows the brightest samples at A♯ while E shows the complete opposite, as it is the parallel key. For C, almost all samples have a bright sentiment.

ACKNOWLEDGMENTS

This work was funded by the Ministry of Science, Research and the Arts Baden-Wuerttemberg in the Artificial Intelligence Software Academy (AISA) and the Cyber Valley Research Fund.

REFERENCES

[1] M. Costa, P. Fine, and P. Ricci Bitti. Interval distributions, mode, and tonal strength of melodies as predictors of perceived emotion. *Music Perception (MP)*, 22:1–14, 2004.

[2] E. Coutinho. *A Neural Network Model for the Prediction of Musical Emotions*. Control, Robotics and Sensors. Institution of Engineering and Technology, 2010.

[3] S. Dalla Bella, I. Peretz, L. Rousseau, and N. Gosselin. A developmental study of the affective value of tempo and mode in music. *Cognition*, 80(3):B1–B10, 2001.

[4] D. Holten and J. J. Van Wijk. Force-directed edge bundling for graph visualization. *Computer Graphics Forum (CGF)*, 28(3):983–990, 2009.

[5] P. N. Juslin and J. A. Sloboda. *Music And Emotion: Theory and research*. Oxford University Press, 08 2001.

[6] M. V. Lakshitha and K. Jayaratne. Melody analysis for prediction of the emotions conveyed by Sinhala songs. In *2016 IEEE International Conference on Information and Automation for Sustainability (ICIAfS)*, pages 1–6, 2016.

[7] A. Micallef Grimaud and T. Eerola. An interactive approach to

emotional expression through musical cues. *Music & Science*, 5:205920432110617, 2022.

[8] R. Panda, B. Rocha, and R. P. Paiva. Music emotion recognition with standard and melodic audio features. *Applied Artificial Intelligence*, 29:313–334, 2015.

[9] D. Ramos, J. Bueno, and E. Bigand. Manipulating greek musical modes and tempo affects perceived musical emotion in musicians and nonmusicians. *Brazilian journal of medical and biological research = Revista brasileira de pesquisas médicas e biológicas / Sociedade Brasileira de Biofísica ... [et al.]*, 44:165–72, 2011.

[10] E. Schellenberg, A. Krysciak, and R. Campbell. Perceiving emotion in melody: Interactive effects of pitch and rhythm. *Music Perception (MP)*, 18:155–171, 2000.

[11] E. Schubert. Modeling perceived emotion with continuous musical features. *Music Perception (MP)*, 21(4):561–585, 2004.