COSINE CONTOURS: A MULTIPURPOSE REPRESENTATION FOR MELODIES

Bas Cornelissen **Willem Zuidema** **John Ashley Burgoyne**
Institute for Logic, Language and Computation, University of Amsterdam
b.j.m.cornelissen@uva.nl, zuidema@uva.nl, j.a.burgoyne@uva.nl

ABSTRACT

Melodic contour is central to our ability to perceive and produce music. We propose to represent melodic contours as a combination of cosine functions, using the discrete cosine transform. The motivation for this approach is twofold: (1) it approximates a maximally informative contour representation (capturing most of the variation in as few dimensions as possible), but (2) it is nevertheless independent of the specifics of the data sets for which it is used. We consider the relation with principal component analysis, which only meets the first of these requirements. Theoretically, the principal components of a repertoire of random walks are known to be cosines. We find, empirically, that the principal components of melodies also closely approximate cosines in multiple musical traditions. We demonstrate the usefulness of the proposed representation by analyzing contours at three levels (complete songs, melodic phrases and melodic motifs) across multiple traditions in three small case studies.

1. INTRODUCTION

Humans are born with a remarkable sensitivity to melodic contour. This is dramatically illustrated when newborns cry: the cries of German babies tend to go down in pitch, but those of French babies go up, even if falling contours are physiologically easier to produce [1]. By imitating the intonation patterns of their mothers' language, babies take the first steps towards a spoken language—helped by exaggerated pitch contours of infant directed speech [2]. Contour perception remains central to speech, for intonation or even word distinctions, but is also a key ingredient of human musicality [3]. Dowling famously argued that melodies are remembered as two independent parts, a scale and a contour [4]. A scale then functions as a ladder "on which the ups and downs of the contour where hung." Indeed, when listening to novel melodies, contours appear to stand out more than the exact intervals and influence the perceived similarity of melodies [5]. That has also motivated studies of contour in MIR, in particular for measuring melodic similarity [6]. As we briefly review below, many representations of contour have been proposed in answer to the recurring question: how can one best describe melodic contour?

We propose representing melodies as combinations of cosine functions. This is motivated by the need for a concise, maximally informative representation: how can we capture as much of the variability in contour data in as few dimensions as possible? The easiest solution would be to use a *principal component analysis* (PCA). In section 4, we show empirically that the principal components of melodies do not take arbitrary shapes, but in fact closely approximate cosines. We then relate this observation to theoretical results showing that the principal components of certain random walks are sinusoidal, as a result of a particular covariance structure. The proposed 'cosine contour' space thus closely approximates the optimal solution provided by PCA, but offers several benefits. The key argument for this representation is theoretical and we leave a systematic comparison of contour representations for future work. Instead we discuss three case studies that demonstrate the usefulness of cosine contours.

Cosine contours meet several desiderata for contour representations. First, a good representation respects the linear structure of melody and is *invariant to transposition and tempo changes*. Second, the representation should be *interpretable* and *intuitive* (and, in particular, avoid some of the shortcomings of polynomial coefficients). Third, the representation should support *variable levels of abstraction*, so that one can interpolate between a broad summary of the shape, and the exact pitch curve. Fourth, we look for a *broadly applicable* and *culturally neutral* representation: it should be able to describe contours from different cultures, or even from different domains (e.g., speech). It should also be able to handle both audio and symbolic data, although we only analyze symbolic data here.

2. WHAT IS MELODIC CONTOUR?

Melodic contour is a general description of a melody's shape that abstracts away from the particular pitches and precise rhythms. It has been characterised in many different ways. Ethnomusicologists (and composers) have used *contour typologies*: small sets of contour types [7]. David Huron, for example, distinguished nine types of contours by comparing the initial and final pitches to the average pitch on the middle part of a melody [8]. When, say, the initial is above the middle, which in turn equals the final,

Attribution: B. Cornelissen, W. Zuidema, and J.A. Burgoyne, "Cosine Contours: A Multipurpose Representation for Melodies", in *Proc. of the 22nd Int. Society for Music Information Retrieval Conf.,* Online, 2021.

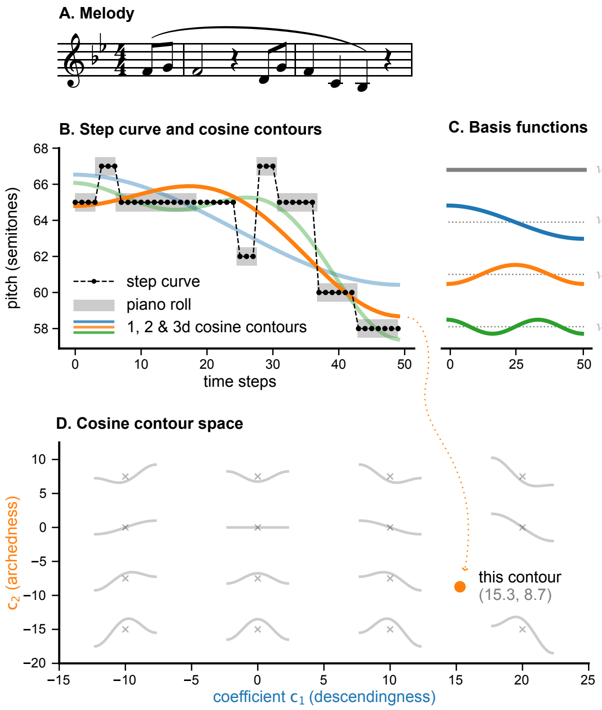

Figure 1. **Cosine contours** represent a melodic contour as a combination of cosine functions. **(A)** This is illustrated for a short melodic phrase. **(B)** A piano roll is interpolated to obtain fixed-length vector of MIDI pitches (black curve). This vector is approximated using a discrete cosine transform (coloured curves). Increasing the dimensionality, from, e.g., 1 (blue) to 3 (green) improves the approximation. **(C)** The basis functions correspond to simple shapes. This makes the cosine contour space interpretable, as illustrated in **(D)** for the first two dimensions. Every point in this space defines a contour shape, varying in what we call the *descendingness* and *archedness*. The orange dot represents the orange contour from **(B)**.

the melody has a ‘descending-horizontal’ contour. Such a formal typology can be used in MIR [9], but typologies have also been defined using verbal descriptions or even drawings [7, 10]. CantoCore, for example, instructs an annotator to look for six types: ascending, descending, arched, U-shaped, undulating and horizontal [11]. Even though the types are less sharply defined, such typologies have inspired cross-cultural generalizations such as the *melodic arch hypothesis*: the claim that melodic phrases tend to be arch-shaped or descending [8, 12–14].

In melody extraction from audio, contours are usually represented by sequences of pitches ordered in time. Various contour features derived from this, such as the range or pitch deviation, have been used in classification tasks [15–18]. Contours in symbolic data can be similarly represented as *step curves* (figure 1B, black line) [19, 20]. *Parsons code* drastically simplifies a step curve [21]. It describes the direction of movement from one note to the next (up, down, or level) and discards interval size and note durations. Variants between these two extremes have also been used, by distinguishing various classes of jump sizes [6]. Another strategy is to focus on salient notes, typically turning points (maxima and minima), and to discard other notes [7, 18, 19]. This often requires special handling of ornaments [20], possibly tailored to the repertoire. Yet another approach considers the relative ordering of all pairs of notes in a melody, summarized in a matrix. Such combinatorial models in way expand rather than reduce the representation, break the linearity of the melody and are sensitive to local changes [20].

Finally, one can describe melodies using continuous functions. Müllensiefen and Wiggins fit a polynomial function to a step curve and use the coefficients to represent the contour [20]. The degree of the polynomial is chosen per phrase, using the Bayesian information criterion (BIC) to avoid overfitting. Polynomial coefficients are quite difficult to interpret, however: they change drastically when the degree changes, and can also be sensitive to changes in the data, especially when the polynomials are not orthogonal and introduce correlations between the coefficients (collinearity). Instead of fitting a function to the contour, one can also *decompose* the contour and express it as a sum of (orthogonal) basis functions. Velarde and colleagues have for example used *Haar wavelets* as basis functions in musical pattern discovery [22]. The step-like shapes of those wavelets are well suited to describe particular melodic patterns, but make them less suited for describing the overall contour. An alternative basis of sinusoidal functions is implicit in Schmuckler’s use of a Fourier analyses to represent melodic contour [23]. This has been interpreted as measuring the ‘periodic information’ in a melody, and was reported to correlate with perceived similarity.

3. DATA

With the broad applicability in mind, we analyze music from several independent traditions. The choice of traditions was partly motivated by our aim to analyze contours at multiple levels of description: we expect (different) regularities at different levels. At the highest level, complete *songs* can have characteristic shapes, and those shapes may differ between traditions. At the smaller level *phrases* may be subject to the melodic arch hypothesis cited above. Finally, at the smallest level, *melodic motifs* could exhibit sequential structure, for example when melodies in a repertoire are formed by stringing together melodic motifs (sometimes called *centonization* [24]). We also analyze *random segments* obtained by slicing a melody at random in approximately phrase-length segments, so that their boundaries usually do not overlap with actual phrase boundaries [25].

One tradition for which all of these levels are directly available is Gregorian chant, thanks to two recently released corpora: the CantusCorpus and the GregoBaseCorpus [25]. Gregorian chant has been sung in Roman Catholic churches for well over a thousand years. The close connection between music and text in chant suggests a natural subdivision of the music into motifs corresponding to words or syllables. The notation suggests even smaller motifs: it is based on small figures, called *neumes*, that represent short groups of notes [26]. To analyse motif contours, we use chants from the CantusCorpus (v0.2) with transcriptions of medieval manuscripts, which include neume boundaries. We focus on the two largest chant genres: *antiphons* and *responsories*. Phrase boundaries are not available in the CantusCorpus, however, and so for that, we turn to the GregoBaseCorpus (v0.3) of modern chant transcriptions. Modern chant notation includes explicit breathing marks (*pausas*), which have been used to extract phrases [25].

Phrase markings are also included in the Essen Folksong Collection [27], from which we analyse phrases from German and Chinese folksongs. We focus on the two largest subsets, 'Erk' [28] (9782 contours) and 'Han' (7601 contours). [1] At the level of complete songs, we also add music from the Sioux people made available in the *Densmore Collection* [29, 30]. In the supplementary material, we include some further analyses of several other traditions from the *Essen* and *Densmore* collections.

We convert all melodies (be it songs, phrases or motifs) to step contours by extracting note onsets (in quarter notes) and pitches (in MIDI semitones). We then interpolate a step function through these points, from which we sample $N = 100$ equally spaced pitches. Those pitches are collected in vectors $\mathbf{x} = (x_0, \ldots, x_{N-1})$ (black curve in figure 1A), which are the basic data analysed in this paper. [2]

Our starting representation makes several assumptions that seem reasonable (and common: [13, 14, 22]) when only interested in contour. First, we ignored all rests. Second, we normalize the duration of all contours. Both 3-note motifs and 30-note songs are represented by vectors of 100 pitches. The relative durations within that melody are of course retained, so we would still see that contours of short motives are probably simpler than those of long melodies. Third, we assume Euclidean distances between melodies. This is usually problematic, but less so when we are only interested in contour similarity. Our analyses require that all contours are embedded in a vector space. Using more sophisticated measures such as dynamic time warping distance, would require us to reconstruct a space (e.g., using multidimensional scaling), and make the analyses less transparent. Finally, note that we do *not* center the contours to have mean pitch 0. This is sometimes done to make contours transposition invariant and more directly comparable [14, 22, 25]. We will soon see that our proposed representation elegantly resolves this problem without requiring centring.

4. PRINCIPAL COMPONENTS OF CONTOURS

In this section, we explore principal component analysis applied to contours. The goal of PCA is to find a set of orthogonal axes, the *principal components*, that contain most of the variance in the dataset. Note that the principal components, like the original contours from our data, are N-dimensional vectors, such that the contours and components can be interpreted and plotted in the same space.

In figure 2A , we show results from applying PCA on a large dataset of Gregorian chant (similar results with German and Chinese folksongs can be found in supplement S2). We plot the first four principal components of several types of melodies: short motifs (syllables), phrases, random segments of melodies, and complete songs. We show responsory syllables from CantusCorpus for the motifs, antiphon phrases from the GregoBaseCorpus and finally all song contours from GregoBaseCorpus.

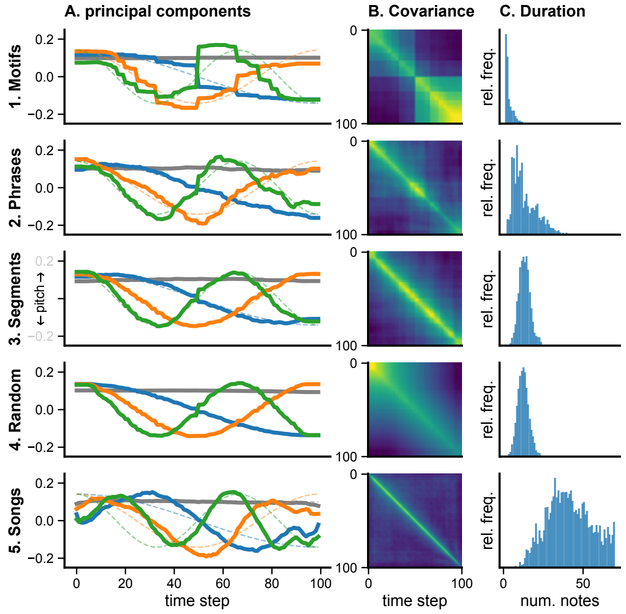

Figure 2. **Principal components of contours** (solid lines) are roughly cosine shaped (dashed) across different levels (**A**). This is a result of the particular structure of the covariance matrix (**B**): matrices of this type have Fourier basis functions as their eigenvectors. This is clearest for phrases (**2**) or random segments from melodies (**3**), here of similar length as phrases. Crucially, we see the same effect for simulated, contour-like random walks (**4**). For complete songs (**5**) the effect is less clear, probably due to differences in typical length (**C**) and data size. Contours in **1–4** are from Gregorian chant.

Surprisingly, we find that the principal components are highly similar across most of those data sets, and correspond to well-known contour shapes: descending, convex, and—perhaps—undulating. This is clearest for the phrases and random segments. For complete songs the effect is weaker, especially for even smaller datasets (see the supplement S2). Besides small data sizes, the fact that songs are much longer also plays a role (see fig. 2C). We also applied the analysis on simulated random walks approximating phrases: we draw the number of notes from a similar length distribution, normalize the duration and then sample $N = 100$ pitches as before (see supplement S1 for details). Interestingly, the pattern is now even clearer, suggesting there must be a mathematical explanation.

To give that explanation, we need to first describe PCA more formally. We consider a collection of M contour vectors $\mathbf{x}_m$ of length N. Denote the sample mean by $\bar{\mathbf{x}} = \frac{1}{M}\sum_m \mathbf{x}_m$ and the centered data by $\hat{\mathbf{x}}_m = \mathbf{x}_m - \bar{\mathbf{x}}$. The first principal component of the dataset is then defined as

[1] Much is unclear about the exact (bibliographic) origins of the Chinese subset of *Essen*. This is problematic given its wide use in computational musicology and deserves further attention from the community.

[2] See github.com/bacor/cosine-contours for data, code and supplements.

a normalized vector $\mathbf{u}_1 \in \mathbb{R}^D$ for which the projected data $\{\mathbf{u}_1^T \mathbf{x}_m : 1 \leq m \leq M\}$ has maximal variance. It can be shown (e.g., [31]) that this is the case when $\mathbf{u}_1$ is an eigenvector corresponding to the largest eigenvalue λ_1 of the covariance matrix

$$\mathbf{S} = \frac{1}{M}\sum_{m=1}^{M}(\mathbf{x}_m - \bar{\mathbf{x}})(\mathbf{x}_m - \bar{\mathbf{x}})^T, \tag{1}$$

so that $\mathbf{S}\mathbf{u}_1 = \lambda_1 \mathbf{u}_1$. It follows that the projected variance is given by λ_1, the largest eigenvalue. The other principal components similarly emerge as the other eigenvectors of the covariance matrix.

The covariance matrices (figure 2B) for both random walks and our empirical data have a particular structure: they *roughly* resemble *Toeplitz matrices*, which have fixed values along each of their diagonals. Such covariance structures are frequently encountered in spatial or temporal data, when the covariance decreases with the distance between the points [32–34]. With the empirical contours that appears to be the case (and for random walks it is there by design): there is higher correlation between successive pitches and lower correlation between distant pitches. As a result, the higher covariances are concentrated along the diagonal. Again, this clearest for the phrases and random segments. For motifs we see some deviations: two 'blocks' in the covariance matrix, and corresponding jumps half way through the principal components. This is easily explained by the fact that motifs often span only two notes. In that case, all pitches in the first half of the contour are then perfectly correlated, as are pitches in the final half. Crucially, despite such deviations from a perfect Toeplitz structure, the principal components are still well-approximated by cosines.

If you let a Toeplitz matrix grow in size, it asymptotically tends towards a *circulant* matrix, preserving properties such as eigenvalues and eigenvectors along the way [32]. Circulant matrices have exactly the same values in every row, but rotated one step to the right with respect to the previous row. This has the surprising result that all circulant matrices have the same eigenvectors: basis vectors of the discrete Fourier transform. For a real and symmetric matrices, like covariance matrices, this results in cosine-shaped eigenvectors of increasing frequency—exactly what we see in figure 2. We discuss all of this in more detail in the supplement S2. In sum, because of a Toeplitz-like covariance structure, the principal components of melodic contours will tend to look like cosine functions.

5. COSINE CONTOURS

Next we turn this observation, and its explanation, into a proposal for a new contour representation. The idea is to approximate the principal components by cosine functions and then project the contours on those first few cosines to obtain a low-dimensional representation. This is exactly equivalent to taking a *discrete cosine transform* (DCT) of the contour [35].

Formally, consider a collection of contours of length N as before. We approximate the k-th principal component $\mathbf{u}_k$ by a vector $\mathbf{v}_k = \big(v_k(0), \ldots, v_k(N-1)\big)$ whose entries are given by the cosine function [3]

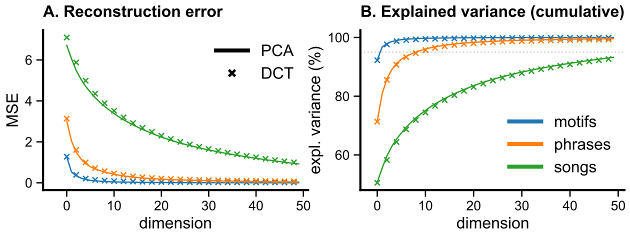

Figure 3. **DCT approximates PCA**, the optimal transform, in terms of the reconstruction error **(A)** and the explained variance ratio **(B)**. The reconstruction error is the mean squared error between an contour and a lower dimensional reconstruction. Note that data corresponds to figure 2, and that we did *not* discard the first component c_0 of the DCT in this figure.

$$v_k(n) = \alpha_k \cdot \cos\frac{\pi(2n+1)k}{2N}. \tag{2}$$

Here $\alpha_0 = 1/\sqrt{N}$ and $\alpha_k = \sqrt{2/N}$ for $k \geq 1$ are normalizing constants ensuring that $\mathbf{v}_k$ has unit norm. The projection of a contour $\mathbf{x} = (x_0, \ldots, x_{N-1})$ on $\mathbf{v}_k$ is then given by the inner product $c_k = \mathbf{v}_k^T \mathbf{x}$. Expanding this gives the usual definition of the discrete cosine transform (DCT-II):

$$c_k = \sum_{n=0}^{N-1} x_n \alpha_k \cos\frac{\pi(2n+1)k}{2N}. \tag{3}$$

Conversely, the contour can be reconstructed from the coefficients c_k using the inverse transform $x_n = \sum_{k=0}^{N-1} c_k v_k(n)$. Using only $D < N$ coefficients, we define our low-dimensional *cosine contour representation* as $C_D(\mathbf{x}) = (c_1, \ldots, c_D)$. Note that we deliberately discard c_0. This coefficient corresponds to a flat line and describes the overall pitch height of a contour: exactly what we need to get rid of to make the contour transposition invariant. In this way we resolve the centering of contours discussed above.

Why use this representation instead of principal components? Indeed, a principal component projection (also known, in this context, as the *Karhunen-Loève transform*), is optimal in several ways [35, 37]. Not only does it decorrelate the data, it also packs most variance in the first few transform coefficients (sometimes called *energy compaction*), and minimizes the reconstruction error when using only a few coefficients. However, the transformation depends on the data. Concretely, the principal components of German phrase contours differ from Chinese ones. Any choice for using one of the two is arbitrary. In contrast, the DCT is a principled, neutral solution—that approximates the optimal transform. In fact, the DCT was originally introduced for similar reasons [35], and was then found to empirically approximates PCA well in domains ranging from image to audio [37]. The current results suggest that the same applies for melodies.

[3] These basis functions correspond to the most popular version of the discrete cosine transform, DCT-II, for which fast implementations are widely available; others would have been possible [36].

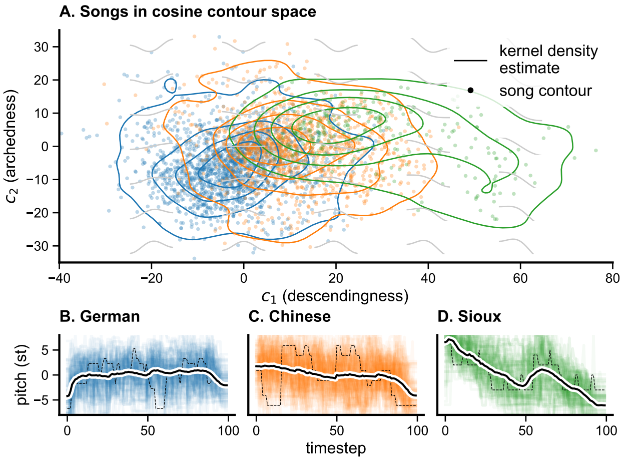

Figure 4. **Songs of three cultures** represented in the cosine contour space (**A**) show substantial variability. The average of all contours in a tradition (**B–D**) also illustrates this (thick black lines; dashed lines highlight one contour).

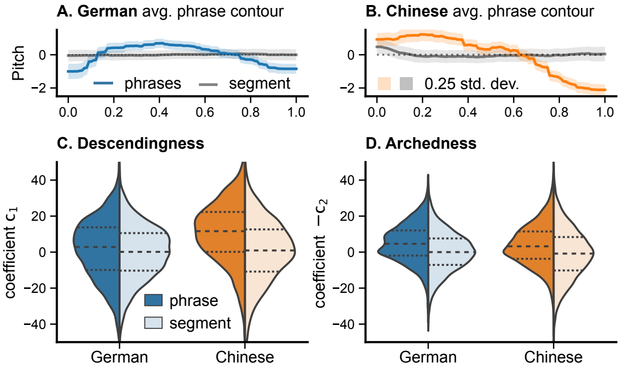

Figure 5. **Phrases** of German (**A**) and Chinese (**B**) songs tend to be more descending and arched compared to random segments from the same melodies, as visible from their average contours. This can be quantified by comparing the first (**C**) and second (**D**) coefficients of their cosine representations.

6. EVALUATION AND CASE STUDIES

We evaluate proposed contour representation by comparing it to a principal component transformation, to demonstrate that representation is close to the optimum. We further designed three case studies to illustrate its usefulness at the levels of (1) song, (2) phrases and (3) motifs. The case studies show that the representation is musicologically meaningful, as it allows visualization of variation (1), a quantitative evaluation of constraints on variation (2), and accurate classification into traditional categories (3). For simplicity, we only look at two dimensional representations in these case studies, but higher dimensions may be useful in practice.

6.1 Optimality

To empirically verify the claim that the DCT approximates the optimal PCA transform, we compute the reconstruction error and the explained variance ratio using the same data as before. The reconstruction error is measured as the mean square error between a contour and its D-dimensional reconstruction, using either the principal components (PCA) or cosines (DCT) as basis functions (so for $D = N$, the reconstruction is guaranteed to be perfect). Figure 3A shows that the reconstruction errors of DCT closely approximate that of PCA. For the shorter contours (motifs and phrases), the error very rapidly decreases, indicating that low-dimensional representations are already effective. Indeed, to explain 95% of the variance using cosine contours, you need 1 dimension for motifs, 9 for phrases and 61 for songs (this is sometimes called the *effective dimensionality* [38]). [4]

6.2 Case Study 1: Visualizing different traditions

Low dimensional representations of song contours are not likely to be very informative, yet we find that some traditions can be somewhat distinguished in just two dimensions. Figure 4 shows song contours from German, Chinese and Sioux songs. Sioux songs have a striking overall shape (subplot D), often strongly descending, which is reflected in the distribution of contour shapes. Similarly, German songs appear to be more arch-like than songs from the other traditions.

6.3 Case Study 2: The melodic arch hypothesis

In a second case study, we look at the melodic arch hypothesis, which states that *phrases* tend to be arch-shaped or descending [8] (see figure 5A, B) in a way that it becomes much easier to test (cf. [14]). We observe that the first component c_1 of a cosine representation roughly measures the *descendingness* of the contour, and, similarly, that $-1 \cdot c_2$ measures the *archedness*. The melodic arch hypothesis can thus be reformulated as stating that c_1 and $-c_2$ are larger for phrases than for random segments of the melodies (cf. [25]). Comparing Chinese and German phrases, we find that all are significantly ($p \ll 0.001$) more descending and arched than the corresponding random segments (see figure 5C, D). This demonstrates that the coefficients of the cosine contour representation are musicologically meaningful.

6.4 Case Study 3: Mode classification

In the final case study, we evaluate the performance of this contour representation on a task: mode classification in plainchant. Gregorian chant uses a system of eight *modes*: Dorian, Phrygian, Lydian and Mixolydian, each in the two flavours plagal and authentic. Modes differ not only in their scales, but also in their melodic movement. Plagal melodies tend to move lower than authentic ones, closer around the tonal center. In a recent paper we suggest that the mode of Gregorian chant can be predicted from contours alone, in that case using a Parsons code contour representation [39]. We sliced up chants in sequences of motifs corresponding to the notational units (so called *neumes*) or textual units: all notes set to one *syllable* of the text would form a unit, and similarly for *words*. Next, we represented chants as

[4] However, note that Moore et al [38] show that high-dimensional random walks can falsely appear to have a low effective dimensionality.

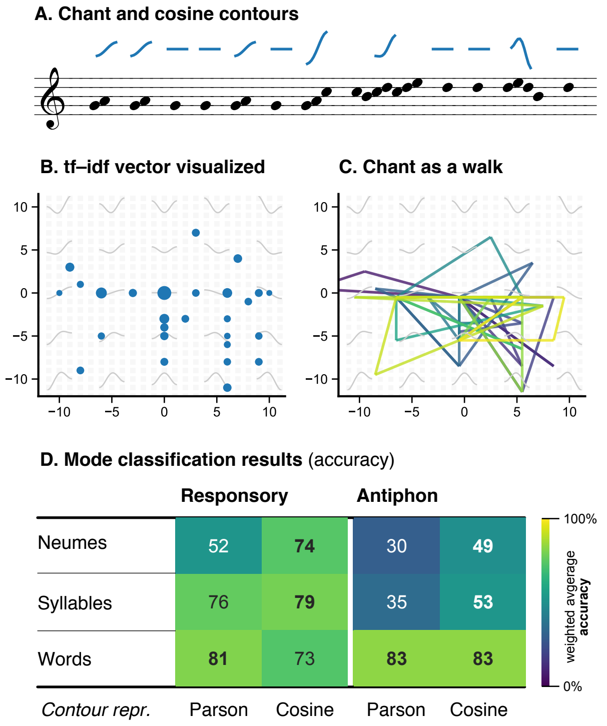

Figure 6. **Motifs used for mode classification** in Gregorian chant. (A) A chant is segmented into motifs derived from the notation (neumes) or lyrics (syllables, words). The blue curves show the two-dimensional cosine contours for those motifs. (B) We discretize the contour space and represent the chant as a vector of tf–idf weighed motif frequencies ('grid cell frequencies'). Dots illustrate the nonzero entries of this vector for the chant shown above. (C) The chant is now a walk through contour space, but our 'bag of motifs' ignores order. (D) Using these vectors to classify mode, we outperform a previous study using a Parsons code for the smaller motifs neumes and syllables.

vectors of motif or *term frequencies* (tf), where each entry was weighted by the *inverse document frequency* (df; the number of chants or documents containing that motif). A linear support vector machine was then trained on these *tf–idf vectors* to predict the mode.

We repeat these experiments using a two-dimensional cosine representation for the motifs rather than a Parsons code. There is one technical problem: whereas cosine contours are continuous, the tf–idf model requires a discrete vocabulary of motifs. We therefore discretize the cosine contour space to a grid, and effectively treat every chant as a sequence of grid-cells (fig. 6C). All in all, this introduces two new parameters to the experiment: the dimensionality of the cosine contour and the resolution of the grid. In this case study, we do not tune these parameters and focus on two dimensional contours, discretized to a grid between −20 and 20 with a grid size of 1. For ease of reading, the figure 6B shows the grid only from −10 to 10.

The results are summarized in figure 6D. We see an interesting pattern: the cosine contours outperform the original results for small motifs such as neumes and syllables, but not for words, which are much longer motifs. This seems to makes sense: two dimensional cosine contours are a fairly crude approximation of those longer contours, but may reasonably approximate short motifs.

7. DISCUSSION AND CONCLUSIONS

This paper proposed a novel representation for melodies using the discrete cosine transform. Observing that the principal components of melodies tend to be shaped like cosines, this representation approximates the optimal representation in the sense that it packs most variance in a few dimensions. First, the cosine representation is easily interpretable, since it presents contours as a linear combination of cosine functions with intuitive shapes. Second, by changing the dimensionality, the level of abstraction of the contour can be varied, allowing arbitrary small reconstruction error by including more and more dimensions. Third, this representation allows one to map contours at multiple levels, from motifs to songs, to one common space. The cosine representation thus creates a common ground for comparing contours across traditions and levels. That is possible as, fourth, the representation is independent of the data, and in that sense culturally neutral.

The observation that principal components of spatial and temporal data can have sinusoidal shapes is not novel, but does not appear to be widely known. Indeed, the sinusoidal shapes have been interpreted as genuine effects, rather than mathematical artefacts. For example, one study interpreted gradients in the principal components of human genetic variation across the world as evidence for certain migration events in human history [40]. Closer inspection revealed that those gradients were sinusoidal 'artefacts' analogous to those reported in the present paper [33]. Closer to MIR, it has been observed that the training trajectories of deep neural networks have sinusoidal principal components [41], for the same reason. Again, a detailed analysis [34] revealed these were artefacts, but accurately reflecting the behaviour of high-dimensional random walks [34, 38]. We hope this paper helps increasing the awareness of this phenomenon.

The present work only begins to explore this new contour representation and raises many further questions. One particularly promising possibility is the application to audio data. We only explored symbolic data, but the proposed representation lends itself well for applications on acoustic data. One application we hope to explore further is the analysis of speech intonation using the cosine contour representation. A possible other avenue would be the analysis of folk song recordings, of which vast collections have been collected. Folk song researchers have often used contour in some way to organize repertoires [7], and this representation may contribute to that. Contour typologies have also be used in cross-cultural comparisons (see e.g. [12]). Many typologies have been proposed [7, 8, 10, 11], but they have not been systematically evaluated, and we think the proposed representation will be valuable there.

8. ACKNOWLEDGEMENTS

We would like to thank Henkjan Honing and Marianne de Heer Kloots for their feedback on the manuscript. We would also like to thank the four anonymous reviewers for their careful reviews; we have tried to address all your comments.

9. REFERENCES

[1] B. Mampe, A. D. Friederici, A. Christophe, and K. Wermke, "Newborns' cry melody is shaped by their native language," *Current Biology*, vol. 19, no. 23, pp. 1994–1997, 2009.

[2] K. Wermke, M. P. Robb, and P. J. Schluter, "Melody complexity of infants' cry and non-cry vocalisations increases across the first six months," *Scientific Reports*, vol. 11, no. 1, p. 4137, Dec. 2021.

[3] H. Honing, C. ten Cate, I. Peretz, and S. E. Trehub, "Without it no music: Cognition, biology and evolution of musicality," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 370, no. 1664, p. 20140088, Mar. 2015.

[4] W. J. Dowling, "Scale and contour: Two components of a theory of memory for melodies." *Psychological Review*, vol. 85, no. 4, pp. 341–354, 1978.

[5] M. A. Schmuckler, "Tonality and contour in melodic processing," in *The Oxford Handbook of Music Psychology*, 2nd ed., S. Hallam, I. Cross, and M. Thaut, Eds. Oxford University Press, 2016.

[6] D. Müllensiefen and K. Frieler, "Cognitive adequacy in the measurement of melodic similarity: Algorithmic vs. human judgments," in *Music Query: Methods, Models, and User Studies*, ser. Computing in Musicology. MIT Press, 2004, no. 13, p. 30.

[7] C. R. Adams, "Melodic contour typology," *Ethnomusicology*, vol. 20, no. 2, p. 179, May 1976.

[8] D. Huron, "The melodic arch in Western folksongs," *Computing in musicology*, vol. 10, pp. 3–23, 1996.

[9] D. Müllensiefen, "Fantastic: Feature ANalysis Technology Accessing STatistics (In a Corpus): Technical Report v1.5," Tech. Rep., 2009.

[10] T. Kelkar, U. Roy, and A. R. Jensenius, "Evaluating a collection of sound-tracing data of melodic phrases," in *19th International Society for Music Information Retrieval Conference*, Paris, France, 2018, pp. 74–81.

[11] P. E. Savage, E. Merritt, T. Rzeszutek, and S. Brown, "CantoCore: A new cross-cultural song classification scheme," in *Analytical Approaches to World Music*, vol. 2, 2012, pp. 87–137.

[12] P. E. Savage, S. Brown, E. Sakai, and T. E. Currie, "Statistical universals reveal the structures and functions of human music," *Proceedings of the National Academy of Sciences*, vol. 112, no. 29, pp. 8987–8992, 2015.

[13] A. T. Tierney, F. A. Russo, and A. D. Patel, "The motor origins of human and avian song structure," *Proceedings of the National Academy of Sciences*, vol. 108, no. 37, pp. 15 510–15 515, Sep. 2011.

[14] P. E. Savage, A. T. Tierney, and A. D. Patel, "Global music recordings support the motor constraint hypothesis for human and avian song contour," *Music Perception: An Interdisciplinary Journal*, vol. 34, no. 3, pp. 327–334, Feb. 2017.

[15] R. M. Bittner, J. Salamon, J. J. Bosch, and J. P. Bello, "Pitch contours as a mid-level representation for music informatics," in *Audio Engineering Society Conference on Semantic Audio*, Erlangen, Germany, 2017.

[16] M. Panteli, R. Bittner, J. P. Bello, and S. Dixon, "Towards the characterization of singing styles in world music," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. New Orleans, LA: IEEE, Mar. 2017, pp. 636–640.

[17] R. M. Bittner, J. Salamon, S. Essid, and J. P. Bello, "Melody extraction by contour classification," in *Proceedings of the 16th International Conference on Music Information Retrieval (ISMIR 2015)*, Malaga, Spain, 2015, pp. 500–506.

[18] J. Salamon, B. Rocha, and E. Gomez, "Musical genre classification using melody features extracted from polyphonic music signals," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Kyoto, Japan: IEEE, Mar. 2012, pp. 81–84.

[19] W. Steinbeck, *Struktur und Ähnlichkeit: Methoden automatisierter Melodieanalyse*. Bärenreiter, 1982.

[20] D. Müllensiefen and G. A. Wiggins, *Polynomial Functions as a Representation of Melodic Phrase Contour*, ser. Hamburger Jahrbuch Für Musikwissenschaft. Peter Lang, Jan. 2012.

[21] D. Parsons, *Directory of tunes and musical themes*. Spencer Brown, 1975.

[22] G. Velarde, D. Meredith, and T. Weyde, "A wavelet-based approach to pattern discovery in melodies," in *Computational Music Analysis*, D. Meredith, Ed. Cham: Springer International Publishing, 2016, pp. 303–333.

[23] M. A. Schmuckler, "Testing models of melodic contour similarity," *Music Perception*, vol. 16, no. 3, pp. 295–326, Apr. 1999.

[24] T. Nuttall, M. G. Casado, V. N. Tarifa, R. C. Repetto, and X. Serra, "Contributing to new musicological theories with computational methods: The case of centonization in Arab-Andalusian music," in *Proceedings of the*

20th International Conference on Music Information Retrieval (ISMIR 2019), Delft, The Netherlands, 2019, pp. 223–228.

[25] B. Cornelissen, W. Zuidema, and J. A. Burgoyne, “Studying large plainchant corpora using Chant21,” in *7th International Conference on Digital Libraries for Musicology*. Montréal QC Canada: ACM, Oct. 2020, pp. 40–44.

[26] T. F. Kelly, “Notation I,” in *The Cambridge History of Medieval Music*. Cambridge University Press, 2018, vol. 1, pp. 236–262.

[27] H. Schaffrath, “The Essen folksong collection in the humdrum kern format,” Center for Computer Assisted Research in the Humanities, Tech. Rep., 1995.

[28] L. Erk and F. M. Böhme, *Deutscher Liederhort: Auswahl der vorzüglicheren deutschen Volkslieder, nach Wort und Weise aus der Vorzeit und Gegenwart*. Leipzig: Breitkopf und Härtel, 1893, vol. 1.

[29] F. Densmore, *Teton Sioux Music*, ser. Bulletin 61 of the Bureau of American Ethnology, Smithsonian Institution. Washington, U.S.A.: Government Printing Office, 1918, no. 61.

[30] D. Shanahan and E. Shanahan, “The Densmore collection of Native American songs: A new corpus for studies of effects of geograpy, language and social function on folk song,” in *Proceedings for the 13th International Conference for Music Perception and Cognition*, Seoul, 2014, pp. 206–208.

[31] I. Jolliffe, *Principal Component Analysis*, 2nd ed., ser. Springer Series in Statistics. Springer, 2002.

[32] R. M. Gray, “Toeplitz and circulant matrices: A review,” *Foundations and Trends in Communications and Information Theory*, vol. 2, no. 3, pp. 155–239, 2006.

[33] J. Novembre and M. Stephens, “Interpreting principal component analyses of spatial population genetic variation,” *Nature Genetics*, vol. 40, no. 5, pp. 646–649, May 2008.

[34] J. M. Antognini and J. Sohl-Dickstein, “PCA of high dimensional random walks with comparison to neural network training,” in *32nd Conference on Neural Information Processing Systems*, Montréal, Canada, Jun. 2018.

[35] N. Ahmed, T. Natarajan, and K. R. Rao, “Discrete cosine transform,” *IEEE Transactions on Computers*, p. 4, 1974.

[36] G. Strang, “The discrete cosine transform,” *SIAM Review*, vol. 41, no. 1, pp. 135–147, 1999.

[37] K. R. Rao and P. C. Yip, *Discrete Cosine Transform: Algorithms, Advantages, Applications*. Boston: Academic Press, 1990.

[38] J. Moore, H. Ahmed, and R. Antia, “High dimensional random walks can appear low dimensional: Application to influenza H3N2 evolution,” *Journal of Theoretical Biology*, vol. 447, pp. 56–64, Jun. 2018.

[39] B. Cornelissen, W. Zuidema, and J. A. Burgoyne, “Mode classification and natural units in plainchant,” in *Proceedings of the 21st International Conference on Music Information Retrieval (ISMIR 2020)*, Montréal, Canada, 2020, p. 869–875.

[40] L. Cavalli-Sforza, P. Menozzi, and A. Piazza, “Demic expansions and human evolution,” *Science*, vol. 259, no. 5095, pp. 639–646, Jan. 1993.

[41] E. Lorch, “Visualizing deep network training trajectories with PCA,” in *Proceedings of the 33rd International Conference on Machine Learning*, New York, U.S.A., 2016, p. 5.