Journal of Mathematics and Music
Vol. 00, No. 00, May 2010, 1–23

RESEARCH ARTICLE

Expectation Arrays: Modelling the Similarity of Pitch Collections

Andrew J. Milne[a*], William A. Sethares[b], Robin Laney[a‡] and David B. Sharp[c§]

[a]*Computing Department, Open University, Milton Keynes, UK*; [b]*Department of Electrical and Computer Engineering, University of Wisconsin, Madison, USA*; [c]*Department of Design, Development, Environment and Materials, Open University, Milton Keynes, UK*

(*d mmmm yyyy; final version received d mmmm yyyy*)

Models of the perceived distance between pairs of pitch collections are a core component of broader models of music cognition. Numerous distance measures have been proposed, including voice-leading [1], psychoacoustic [2, 3], and pitch and interval class distances [4]; but, so far, there has been no attempt to bind these different measures into a single mathematical or conceptual framework, nor to incorporate the uncertain or probabilistic nature of pitch perception.

This paper embeds pitch collections in multi-way *expectation arrays* and shows how metrics between such arrays can model their perceived dissimilarity. Expectation arrays indicate the expected number of tones, ordered pairs of tones, ordered triples of tones, etc., that are heard as having any given pitch, dyad of pitches, triad of pitches, etc.. The pitches can be either absolute or relative (in which case the arrays are invariant with respect to transposition). Examples are given to show how the metrics are in accord with musical intuition.

Keywords: music cognition; tone; tonality; microtonality; pitch; salience; expectation; expectation array; metric

MCS/CCS/AMS Classification/CR Category numbers: 05A05; 05A10; 60C05

*Corresponding author. Email: andymilne@tonalcentre.org

ISSN: 1745-9737 print/ISSN 1745-9745 online

1. Introduction

A *pitch collection* may comprise the pitches of tones in a chord, a scale, a tuning, or the virtual and spectral pitches heard in response to complex tones or chords. Modelling the perceived distance (the similarity or dissimilarity) between pairs of pitch collections has a number of important applications in music analysis and composition, in modelling of musical cognition, and in the design of musical tunings. For example, voice-leading distances model the overall distance between two chords as a function of the pitch distance moved by each voice (see [1] for a survey); musical set theory considers the similarities between the interval (or triad, tetrad, etc.) contents of pitch collections (see [4] for a survey); psychoacoustic models of chordal distance [2, 3] treat tones or chords as collections of virtual and spectral pitches [5, 6] to determine their affinity; tuning theory requires measures that can determine the distance between scale tunings and, notably, the extent to which different scale tunings can approximate privileged tunings of intervals or chords (e.g., just intonation intervals with frequency ratios such as $3/2$ and $5/4$, or chords with frequency ratios such as $4:5:6:7$).

This paper presents a novel family of embeddings (*expectation arrays*), and associated metrics, that can be applied to the above areas. Expectation arrays model the uncertainties of pitch perception by "smearing" each pitch over a range of possible values, and the width of the smearing can be derived directly from experimentally determined frequency difference limens [7]. The arrays can embed either absolute or relative pitches (denoted *absolute* and *relative expectation arrays*, respectively): in the latter case, embeddings of pitch collections that differ only by transposition have zero distance; a useful feature that relates similarity to structure.

To avoid confusion, it is worth making some definitions explicit. A *tone* is defined as any stimulus capable of producing a perception of *pitch*. The probability of hearing a tone or a specific pitch is, following Parncutt [2], denoted *salience* (the context should make clear whether salience refers to a tone or a pitch). Two assumptions are made to simplify the analysis: any given tone can be heard as having no more than one pitch and the hearing (or not) of a tone does not affect the chance of hearing another tone. Thus a single note played by an instrument can still be treated as a single perceptual entity or as a set of virtual or spectral "tones". *Pitch collections* are treated as multisets—duplication of the same pitch is meaningful because two different tones may induce the same pitch while both remain discriminable.

2. Category domain embeddings

There are some circumstances where each tone in a pitch collection can be meaningfully related to a unique tone in another collection. This occurs when there are d categories in each pitch collection; for example, categorisation may be by voice (e.g., bass, tenor, alto, soprano), ordinal position within a scale (the scale degree), or the ordinal or metrical position within a theme or melody. When such categorisations occur, pitches may be embedded into a *category domain vector* where the position (index) indicates category and the value indicates pitch. Any standard metric applied to two such category domain vectors provides a pairwise comparison between the pitches of tones in the same category. For example, Chalmers [9] measures the distances between differently tuned tetrachords using a variety of metrics such as Euclidean ℓ_2, taxicab, ℓ_1, and max-value ℓ_∞, and the

Table 1. These pc-vectors represent several musical scales with $b = 2$ (the octave) and $q = 1200$ cents: 12 equal division of the octave (12-EDO), the major scale in 12-EDO, 10-EDO, and a just intonation major scale.

12-EDO	$(0, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100)$	$\mathbb{R}^{12}$
Maj-12	$(0, 200, 400, 500, 700, 900, 1100)$	$\mathbb{R}^{7}$
10-EDO	$(0, 120, 240, 360, 480, 600, 720, 840, 960, 1080)$	$\mathbb{R}^{10}$
Maj-JI	$(0, 204, 386, 498, 702, 884, 1088)$	$\mathbb{R}^{7}$

use of various metrics to measure voice-leading distance are discussed by Tymoczko [1].

To be concrete, a *pitch vector* $\boldsymbol{x}_{\mathrm{pi}} \in \mathbb{R}^d$ contains elements x_{pi_i} indexed by $i \in \mathbb{N} : 1 \leq i \leq d$, where $d \in \mathbb{N}$ is the number of tones. The index i indicates the tone category and the value of the element x_{pi_i} indicates pitch. A typical example is a logarithmic function of frequency

$$x_{\mathrm{pi}_i} = q \log_b \left(\frac{f_i}{f_{\mathrm{ref}}} \right), \tag{1}$$

where $0 < b \in \mathbb{R}$ is the frequency ratio of the period (typically the octave, so $b = 2$), $q \in \mathbb{N}$ determines the number of *pitch units* that make up the period (typically $q = 12$ semitones or $q = 1200$ cents), $f_i \in \mathbb{R}$ is the frequency of tone i, and $f_{\mathrm{ref}} \in \mathbb{R}$ is the frequency given a pitch value of zero (typically C_{-1}, which is 69 semitones below concert A, so $f_{\mathrm{ref}} = 440 \times 2^{-69/12} \approx 8.176$ Hz). With these constants, a four-voice major triad in close position with its root on middle C is $(60, 64, 67, 72)$, which is also the MIDI note numbers for a C major chord.

A *pitch class vector* or *pc-vector*,

$$x_{\mathrm{pc}_i} = x_{\mathrm{pi}_i} \pmod{q}, \tag{2}$$

is invariant with respect to the period of the pitches since $0 \leq x_{\mathrm{pc}_i} \leq q - 1$. This makes it useful for concisely describing periodic pitch collections, such as scales or tunings that repeat every octave. The variable f_{ref} specifies which pitch class has a value of 0 (in a tonal context, it may be clearest to make it equal to the pitch of the root, or tonic). For example, a major triad may be notated $(0, 4, 7)$ or $(1, 5, 8)$, or more generally as $(x, 4 + x, 7 + x) \bmod q$. Table 1 shows some musical scales represented as pc-vectors.

The pc-vector may have an associated *weighting vector*,

$$\boldsymbol{x}_{\mathrm{w}} \in \mathbb{R}^d, \tag{3}$$

which contains elements $0 \leq x_{\mathrm{w}_i} \leq 1$. This can be used to represent amplitude, loudness, salience, and so forth. This paper assumes the weighting vector denotes salience, the probability of hearing a tone. For example, if four voices sound the pitch classes $(0, 3, 3, 7)$ and have an associated weighting vector $(.9, .6, .6, .9)$, listeners are expected to hear the pitch of an outer tone in nine out of ten trials and the pitch of an inner tone in six out of ten trials.

Category domain embeddings, and metrics reliant upon them, are unsuitable when the pitches cannot be uniquely categorised. For example, when modelling the distance between the large sets of spectral or virtual pitches heard in response to complex tones

or chords (see Ex. 6.2), there is no unique way to reasonably align each spectral pitch of one complex tone or chord with each spectral pitch of another [8] and, even if there were, it is not realistic to expect humans to track the "movements" of such a multitude of pitches.

A simpler example is provided by the scales in Table 1, where the categories are the indices of the scale elements. From a musical perspective, it is clear that some such tunings can be thought of as closer than others. For instance, a piece written in Maj-JI can be played in a subset of 12-EDO (such as Maj-12) without undue strain, yet may not be particularly easy to perform when the pitches are translated to a subset of 10-EDO. Thus it is desirable to have a metric that allows a statement such as "Maj-JI is closer to 12-EDO than to 10-EDO."

When two pc-vectors have the same number of elements, any reasonable metric can be used to describe the distance between them; for example, the distance between Maj-12 and Maj-JI can be easily calculated because they both contain seven pitch classes. However, when two pitch collections have different cardinalities, there is no obvious way to define a metric since this would require a direct comparison of elements in $\mathbb{R}^n$ with elements in $\mathbb{R}^m$ for $n \neq m$. One strategy is to identify subsets of the elements of the pitch collections and then try to calculate a distance in this reduced space. For instance, one might attempt to calculate the distance between Maj-JI and 12-EDO by first identifying the seven nearest elements of the 12-EDO scale, and then calculating the distance in $\mathbb{R}^7$. Besides the obvious problems with identifying corresponding tones in ambiguous situations, the triangle inequality will fail in such schemes. For example, let pitch collection $\boldsymbol{x}$ be 12-EDO, pitch collection $\boldsymbol{y}$ be any seven note subset drawn from 12-EDO (such as the major scale), and pitch collection $\boldsymbol{z}$ be a different seven note subset of 12-EDO. The identification of pitches is clear since $\boldsymbol{y}$ and $\boldsymbol{z}$ are subsets of $\boldsymbol{x}$. The distances $d\big(\boldsymbol{x},\boldsymbol{y}\big)$ and $d\big(\boldsymbol{x},\boldsymbol{z}\big)$ are zero under any reasonable metric since $\boldsymbol{y} \subset \boldsymbol{x}$ and $\boldsymbol{z} \subset \boldsymbol{x}$, yet $d\big(\boldsymbol{y},\boldsymbol{z}\big)$ is non-zero because the pitch classes in the two scales are not the same. Hence the triangle inequality $d\big(\boldsymbol{y},\boldsymbol{z}\big) \leq d\big(\boldsymbol{y},\boldsymbol{x}\big) + d\big(\boldsymbol{x},\boldsymbol{z}\big)$ is violated. Analogous counter-examples can be constructed whenever $n \neq m$.

3. Pitch domain embeddings

A way to compare pitch collections with differing numbers of elements is use a *pitch domain embedding* where the index represents pitch and the value represents the probability of a pitch being heard, or the expected number of tones heard at that pitch. Because the cardinality of the pitch domain embedding is independent of the cardinality of the pc-vector they are derived from, such embeddings (and metrics reliant upon them) are able to compare pitch collections with different numbers of tones such as the spectral and virtual pitches heard in response to a complex tone or chord, or scales and their tunings. The following examples are shown as transformations of pc-vectors (2), but they can also be given in terms of pitch vectors (1).

A pc-vector $\boldsymbol{x}_{\mathrm{pc}}$ can be transformed into a characteristic function and weighted by its salience vector $\boldsymbol{x}_{\mathrm{w}}$. The d row vectors are then arranged into a $d \times q$ matrix to allow the saliences of the voices to be individually convolved and appropriately summed. Formally, the elements of the *pitch class salience matrix* $\mathbf{X}_{\mathrm{pcs}} \in \mathbb{R}^{d \times q}$ are given by

$$x_{\mathrm{pcs}_{i,j}} = x_{\mathrm{w}_i}\, \delta(j - \lfloor x_{\mathrm{pc}_i} \rceil), \tag{4}$$

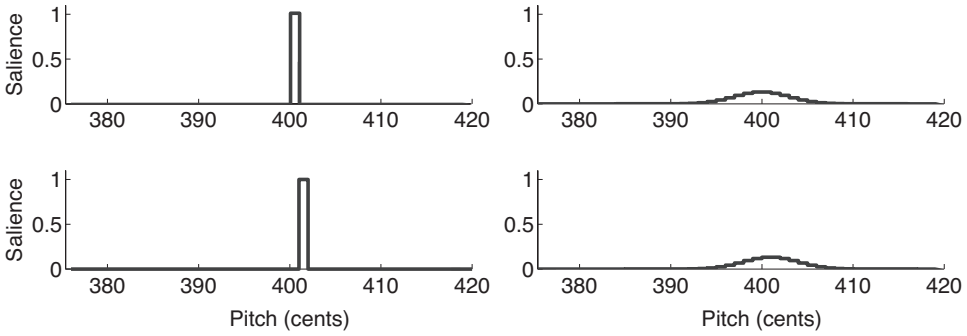

Figure 1. Pitch domain embeddings of two tones—one with a pitch of 400 cents, the other with a pitch of 401 cents. On the left, no smoothing is applied, so their distance under any standard metric is maximal; on the right, Gaussian smoothing (standard deviation of 3 cents) is applied, so their distance under any standard metric is small.

where $[x]$ rounds x to the nearest integer and $\delta(k)$ is the Kronecker delta function that is 1 when $k = 0$ and 0 for all $k \neq 0$.

Example 3.1 Given $q = 12$, $\boldsymbol{x}_{\text{pc}} = (0, 3, 3, 7)$ (i.e., a close position minor chord with a doubled third), and $\boldsymbol{x}_{\text{w}} = (1, .6, .6, 1)$, (4) gives the pitch class salience matrix $\mathbf{X}_{\text{pcs}} = \begin{pmatrix} 1&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&.6&0&0&0&0&0&0&0&0 \\ 0&0&0&.6&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&1&0&0&0&0 \end{pmatrix}$.

Pitch values in the pc-vector are rounded to the nearest pitch unit (whose size is determined by q and b) when embedded in the pitch domain. Using a low value of q (like 12 in the Example 3.1) makes such pitch domain embeddings insensitive to the small changes in tuning that are important when exploring the distances between differently tuned scales, or between collections of virtual and spectral pitches. Embedding into a more finely grained pitch domain (such as $q = 1200$) must be done with care. For example, under any standard metric, the distance between a tone with a pitch of 400 cents and a tone with a pitch of 401 cents is maximally large (i.e., there is no pair of pitches that will produce a greater distance, see the left side of Figure 1). This is counter to perception since it is likely that two such tones will be heard as having pitches that are identical.

The solution is to smooth each spike over a range of pitches to account for perceptual inaccuracies and uncertainties. Indeed, a central tenet of signal detection theory [11] is that a stimulus produces an internal (perceptual) response that may be characterised as consisting of both signal plus noise. The noise component is typically assumed to have a Gaussian distribution, so the internal response to a specific frequency may be modelled as a Gaussian centred on that frequency. It is this noise component that makes the frequency difference limen greater than zero: when two tones of similar, but non-identical, frequency are played successively, the listener may, incorrectly, hear them as having the same pitch. The right side of Figure 1, for instance, shows the effect of smoothing with a Gaussian kernel with a standard deviation of 3 cents. See Appendix A for a detailed discussion of this parameter.

The smoothing is achieved by convolving each row vector in the pitch class salience matrix $\mathbf{X}_{\text{pcs}}$ with a probability mass function. The *pitch class response matrix* $\mathbf{X}_{\text{pcr}} \in$

$\mathbb{R}^{d\times q}$ is given by

$$x_{\text{pcr}_{i,j}} = x_{\text{pcs}_{i,j}} * p_j \tag{5}$$

where p_j is a discrete probability mass function (i.e., $p_j \geq 0$ and $\sum p_j = 1$), and $*$ is convolution (circular over the period q when a pc-vector is used). The result of (5) is that each Kronecker delta spike in $\mathbf{X}_{\text{pcs}}$ is smeared by the shape of the probability mass function and scaled so the sum of all its elements is the salience of the voice.

Example 3.2 Let the probability mass function be triangular with a full width at half maximum of two semitones; this is substantially less accurate than human pitch perception and a much finer pitch granulation (like cents) would ordinarily be required, but it illustrates the mathematics. Applying this to the pitch class salience matrix of Example 3.1 gives the pitch class response matrix $\mathbf{X}_{\text{pcr}} = \begin{pmatrix} .5 & .25 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .25 \\ 0 & 0 & .15 & .3 & .15 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & .15 & .3 & .15 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & .25 & .5 & .25 & 0 & 0 & 0 \end{pmatrix}$.

4. Expectation arrays

The values in the pitch class response matrix represent probabilities; this means it is possible to derive two useful types of array embeddings: (a) *expectation arrays* indicate the expected number of tones, ordered pairs of tones, ordered triples of tones, and so forth, that will be heard as having any given pitch, dyad of pitches, triad of pitches, and so forth; and (b) *salience arrays* indicate the salience of any given pitch, dyad of pitches, triad of pitches, and so forth.

Example 3.2 will help to clarify this distinction: The *expected* number of tones heard at pitch class 3 is 0.6 (the sum of elements with $j = 3$); this does not mean it is possible to hear a non-integer number of tones, it means that over a large number of "trials" an average of 0.6 tones will be heard at pitch class 3 (e.g., given one hundred trials, listeners might hear two tones at pitch class 3 in nine trials, one tone at pitch 3 in forty two trials, and hear no tones at pitch 3 in forty nine trials). The *salience* (probability of hearing) a pitch class of 3 is $1-((1-0)(1-.3)(1-.3)(1-0)) = .51$ so, given one hundred trials, we expect listeners to hear pitch class 3 a total of fifty-one times (regardless of the number of tones heard at that pitch). This paper focuses on expectation arrays.

Expectation arrays may be absolute or relative: *absolute expectation arrays*, denoted $\mathbf{X}_{\text{e}}$, distinguish pitch collections that differ by transposition (e.g., the scales C major and D major), while *relative expectation arrays*, denoted $\hat{\mathbf{X}}_{\text{e}}$, do not.

Expectation arrays enable different pitch collections to be compared according to their monad (single pitch), dyad, triad, tetrad, and so forth, content. To see why such comparisons are significant, consider a simple example using major and minor triads $(0, 4, 7)$ and $(0, 3, 7)$ with $q = 12$. These contain the same set of intervals (and hence they have zero dyadic distance) but these intervals are arranged in different ways (and hence have non-zero triadic distance). Thus the two types of embedding may capture the way major and minor triads are heard to be simultaneously similar and different. MATLAB routines were used to calculate the arrays discussed below; they can be downloaded from `http://eceserv0.ece.wisc.edu/~sethares/pitchmetrics.html`.

4.1. Monad expectation arrays

The *absolute monad expectation vector* $\mathbf{X}_{\mathrm{e}}^{(1)}$ indicates the expected number of tones that will be heard as corresponding to each possible pitch (class) j. It useful for comparing the similarity of pitch collections where absolute pitch is meaningful; for example, comparing the spectral or virtual pitches produced by two complex tones or chords in order to determine their affinity or fit (see Ex. 6.2). The elements of $\mathbf{X}_{\mathrm{e}}^{(1)}$ are

$$x_{\mathrm{e}_j} = \sum_{i=1}^{d} x_{\mathrm{pcr}_{i,j}}, \tag{6}$$

which is equivalent to the column sum of $\mathbf{X}_{\mathrm{pcr}}$. Applied to Example 3.2, (6) produces $\mathbf{X}_{\mathrm{e}}^{(1)} = (0.5, 0.25, 0.3, 0.6, 0.3, 0, 0.25, 0.5, 0.25, 0, 0, 0.25)$.

When there is no probabilistic smoothing, and every voice has a salience of 1, the monad expectation vector is equivalent to a multiplicity function of the rounded pitch (class) vector; that is, $x_{\mathrm{e}_j} = \sum_{i=1}^{d} \delta(j - \lfloor x_{\mathrm{pc}_i} \rceil)$. For example, given the pitch class vector for a four-voice minor triad with a doubled third $(0, 3, 3, 7)$, a weighting vector of $(1, 1, 1, 1)$, and no smoothing, the resulting absolute monad expectation vector is $\mathbf{X}_{\mathrm{e}}^{(1)} = (1, 0, 0, 2, 0, 0, 0, 1, 0, 0, 0, 0)$.

The *relative monad expectation scalar* $\hat{\mathbf{X}}_{\mathrm{e}}^{(0)}$ gives the overall number of tones that will be heard (at any pitch). It can be calculated by summing $\mathbf{X}_{\mathrm{e}}^{(1)}$ over j or, more straightforwardly, as the sum of the elements of the weighting vector

$$\hat{x}_{\mathrm{e}} = \sum_{j=0}^{q-1} x_{\mathrm{e}_j} = \sum_{i=1}^{d} x_{\mathrm{w}_i} \tag{7}$$

Applied to Example 3.2, (7) gives $\hat{\mathbf{X}}_{\mathrm{e}}^{(0)} = 3.2$.

4.2. Dyad expectation arrays

The *absolute dyad expectation matrix* $\mathbf{X}_{\mathrm{e}}^{(2)}$ indicates the expected number of tone pairs that will be heard as corresponding to any given dyad of absolute pitches. It is useful for comparing the absolute dyadic structures of two pitch collections; for example, to compare scales according to the number of dyads they share—the scales C major and F major contain many common dyads and so have a small distance (.1548), the scales C major and F♯ major contain just one common dyad {B, F} and so have a large distance (.7818). (These distances are calculated with a cosine metric (17) and $q = 12$.)

For the dyad arrays, and the higher-dimensional arrays discussed subsequently, an additional family of indices is required: $k_2, k_3, \ldots, k_r$ indicate the pitch of tones relative to a specified pitch j. Thus j, k_2, k_3 identifies a pitch collection with the pitches j, $j + k_2$, and $j + k_3$.

Given two tones indexed by 1 and 2, there are two ordered pairs $(1, 2)$ and $(2, 1)$; the probability of hearing tone 1 as having pitch j and tone 2 as having pitch $j + k_2$

is given by $x_{\text{pcr}_{1,j}} x_{\text{pcr}_{2,j+k_2}}$. Similarly, the probability of hearing tone 2 as having pitch j and tone 1 as having pitch $j+k_2$ is given by $x_{\text{pcr}_{2,j}} x_{\text{pcr}_{1,j+k_2}}$. Given two tones, the expected number of ordered tone pairs that will be heard as having pitches j and $j+k_2$ is, therefore, given by $x_{\text{pcr}_{1,j}} x_{\text{pcr}_{2,j+k_2}} + x_{\text{pcr}_{2,j}} x_{\text{pcr}_{1,j+k_2}}$.

Similarly, given three tones indexed by 1, 2, and 3, there are six ordered pairs $(1,2)$, $(1,3)$, $(2,1)$, $(2,3)$, $(3,1)$, and $(3,2)$; the probability of hearing each pair as having pitches j and $j+k_2$, respectively, are $x_{\text{pcr}_{1,j}} x_{\text{pcr}_{2,j+k_2}}$, $x_{\text{pcr}_{1,j}} x_{\text{pcr}_{3,j+k_2}}$, $x_{\text{pcr}_{2,j}} x_{\text{pcr}_{1,j+k_2}}$, $x_{\text{pcr}_{2,j}} x_{\text{pcr}_{3,j+k_2}}$, $x_{\text{pcr}_{3,j}} x_{\text{pcr}_{1,j+k_2}}$, $x_{\text{pcr}_{3,j}} x_{\text{pcr}_{2,j+k_2}}$. Given three tones, the expected number of ordered tone pairs heard as having pitches j and $j+k_2$ is given by the sum of the above probabilities.

Generalising for any number of tones, the absolute dyad expectation matrix, $\mathbf{X}_{\text{e}}^{(2)} \in \mathbb{R}^{q \times q}$, contains elements

$$x_{\text{e}_{j,k_2}} = \sum_{\substack{(i_1,i_2) \in D^2: \\ i_1 \neq i_2}} x_{\text{pcr}_{i_1,j}} \; x_{\text{pcr}_{i_2,j+k_2}} \,, \tag{8}$$

where $D = \{1, 2, \ldots, d\}$. Element indices j and k_2 indicate the pitches j and $j+k_2$. The element value indicates the expected number of ordered pairs of tones heard as having those pitches.

Equation (8) requires $O(d^2)$ operations for each element. Letting $\mathbf{X}_k$ represent the kth column of the pitch class response matrix $\mathbf{X}_{\text{pcr}}$ and $\mathbf{1}'_d \in \mathbb{R}^d$ be the vector of all ones, this can be simplified to $O(d)$ using Lemma B.1 to

$$x_{\text{e}_{j,k_2}} = \left(\mathbf{1}'_d \mathbf{X}_j\right) \left(\mathbf{1}'_d \mathbf{X}_{j+k_2}\right) - \mathbf{X}'_j \mathbf{X}_{j+k_2}. \tag{9}$$

For example, given the pitch class vector for a four-voice minor triad with a doubled third $(0,3,3,7)$ and a weighting vector of $(1,1,1,1)$, the resulting absolute dyad expectation matrix is $\mathbf{X}_{\text{e}}^{(2)} = \begin{pmatrix} 0&0&0&2&0&0&0&1&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \\ 2&0&0&0&2&0&0&0&0&2&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&1&0&0&2&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0 \end{pmatrix}$. This example is indexed from top to bottom by $j = 0, 1, \ldots, 11$, and from left to right by $k_2 = 0, 1, \ldots, 11$. The first row shows there are two ordered pairs of tones containing the dyad of pitches $\{0,3\}$ (ordered tone pairs $(1,2)$ and $(1,3)$); and one ordered tone pair comprising the dyad of pitches $\{0,7\}$ (tone pair $(1,4)$). Similarly, row 4 shows there are two ordered tone pairs containing the dyad of pitches $\{3,3\}$ (tone pairs $(2,3)$ and $(3,2)$); two ordered tone pairs containing the dyad of pitches $\{3,7\}$ (tone pairs $(2,4)$ and $(3,4)$); two ordered pairs containing the dyad of pitches $\{3,0\}$ (tone pairs $(2,1)$ and $(3,1)$). And so forth.

The *relative dyad expectation vector* $\hat{\mathbf{X}}_{\text{e}}^{(1)}$ indicates the expected number of tone pairs that will be heard as corresponding to any given dyad of relative pitches. It is useful for comparing the intervallic structures of two or more pitch collections regardless of transposition. For example, to compare the number of intervals that two pitch collections have in common or to compare different pitch collections by the number, and tuning accuracy, of a specific set of privileged intervals they each contain (for a specific application, see Example 6.4, which compares thousands of scale tunings to a set of just intonation

intervals).

Summing $\mathbf{X}_{\mathrm{e}}^{(2)}$ over j gives the relative dyad expectation vector $\hat{\mathbf{X}}_{\mathrm{e}}^{(1)} \in \mathbb{R}^q$ with elements $\hat{x}_{\mathrm{e}_{k_2}}$ indexed by $0 \leq k_2 \leq q-1$, where the index indicates interval class:

$$\hat{x}_{\mathrm{e}_{k_2}} = \sum_j x_{\mathrm{e}_{j,k_2}} \tag{10}$$

Assuming the independence of voice saliences, the values are the expected number of ordered tone pairs heard as having that interval, regardless of transposition.

When there is no probabilistic smoothing applied, and the salience of every tone is 1, the relative dyad expectation vector simply gives the multiplicity of ordered pairs of tones that correspond to any possible interval size. For instance, given the pitch class vector for a four-voice minor triad with a doubled third $(0, 3, 3, 7)$ and a weighting vector of $(1, 1, 1, 1)$, the resulting relative dyad expectation vector is $\hat{\mathbf{X}}_{\mathrm{e}}^{(1)} = (2, 0, 0, 2, 2, 1, 0, 1, 2, 2, 0, 0)$. The elements of this vector show that this chord voicing contains 2 ordered pairs of tones with sizes of zero semitones (tone pairs $(2, 3)$ and $(3, 2)$), no ordered pairs of tones with a size of one semitone, no ordered pairs of tones with a size of two semitones, 2 ordered pairs of tones with sizes of three semitones (tone pairs $(1, 2)$ and $(1, 3)$), 2 ordered pairs of tones with sizes of four semitones (tone pairs $(2, 4)$ and $(3, 4)$), and so forth.

When there are no tones with the same pitch class (this is always the case, by definition, when using a pitch class set rather than a multiset), the zeroth element of the interval class vector always has a value of 0. Because the values of all its elements are symmetrical about the zeroth element, no information is lost by choosing the subset $\left\{\hat{x}_{\mathrm{e}_{k_2}} : 1 \leq k_2 \leq \lfloor \frac{q}{2} \rfloor\right\}$ and, when q is an even number, dividing the last element by two (otherwise it is double-counted). When $q = 12$, this subset is identical to the *interval vector* of atonal music theory [10]. The relative dyad expectation array can, therefore, be thought of as a generalisation of a standard interval vector: generalised in that it can deal meaningfully with doubled pitches and the uncertainties of pitch perception.

4.3. Triad expectation arrays

The *absolute triad expectation array* $\mathbf{X}_{\mathrm{e}}^{(3)}$ indicates the expected number of ordered tone triples that will be heard as corresponding to any given triad of absolute pitches. It is useful for comparing the absolute triadic structures of two pitch collections; for example, to compare two scales according to the number of triads they share—the scales C major and F major have many triads in common (e.g., {C, E, G}, {C, D, E}, and {D, F, G}) and so have a small distance (.1702), the scales C major and F♯ major have no triads in common—they share only two notes {B, F}—and so have the maximal distance of 1. (These distances are calculated with the generalised cosine metric (17) with $q = 12$.)

Given three tones indexed by 1, 2, and 3, there are six ordered triples $(1, 2, 3)$, $(2, 1, 3)$, $(2, 3, 1)$, $(1, 3, 2)$, $(3, 1, 2)$, $(3, 2, 1)$; the probabilities of hearing each triple as having pitches j, $j + k_2$ and $j + k_3$, respectively, are $x_{\mathrm{pcr}_{1,j}} x_{\mathrm{pcr}_{2,j+k_2}} x_{\mathrm{pcr}_{3,j+k_3}}$, $x_{\mathrm{pcr}_{2,j}} x_{\mathrm{pcr}_{1,j+k_2}} x_{\mathrm{pcr}_{3,j+k_3}}$, $x_{\mathrm{pcr}_{2,j}} x_{\mathrm{pcr}_{3,j+k_2}} x_{\mathrm{pcr}_{1,j+k_3}}$, $x_{\mathrm{pcr}_{1,j}} x_{\mathrm{pcr}_{3,j+k_2}} x_{\mathrm{pcr}_{2,j+k_3}}$, $x_{\mathrm{pcr}_{3,j}} x_{\mathrm{pcr}_{1,j+k_2}} x_{\mathrm{pcr}_{2,j+k_3}}$, and $x_{\mathrm{pcr}_{3,j}} x_{\mathrm{pcr}_{2,j+k_2}} x_{\mathrm{pcr}_{1,j+k_3}}$. Given three tones, the expected number of ordered tone

triples heard as having pitches $j, j+k_2, j+k_3$ is given by the sum of the above probabilities.

Generalising for any number of tones, the absolute triad expectation array, $\mathbf{X}_{\mathrm{e}}^{(3)} \in \mathbb{R}^{q\times q\times q}$ contains elements

$$x_{\mathrm{e}_{j,k_2,k_3}} = \sum_{\substack{(i_1,i_2,i_3)\in D^3:\\ i_1\neq i_2, i_1\neq i_3, i_2\neq i_3}} x_{\mathrm{pcr}_{i_1,j}}\ x_{\mathrm{pcr}_{i_2,j+k_2}}\ x_{\mathrm{pcr}_{i_3,j+k_3}} \tag{11}$$

where $D = \{1, 2, \ldots, d\}$. Element indices j, k_2, and k_3 indicate the pitch (classes) j, $j+k_2$, and $j+k_3$; assuming the independence of voice saliences, element value indicates the expected number of ordered triples of tones heard as having those three pitches.

Equation (11) requires $O(d^3)$ operations for each element, but can be simplified as in Lemma B.2 to

$$\begin{aligned} x_{\mathrm{e}_{j,k_2,k_3}} &= \left(\mathbf{1}_d'\mathbf{X}_j\right)\left(\mathbf{1}_d'\mathbf{X}_{j+k_2}\right)\left(\mathbf{1}_d'\mathbf{X}_{j+k_3}\right) - \left(\mathbf{1}_d'\mathbf{X}_{j+k_3}\right)\mathbf{X}_j'\mathbf{X}_{j+k_2} \\ &\quad - \left(\mathbf{1}_d'\mathbf{X}_{j+k_2}\right)\mathbf{X}_j'\mathbf{X}_{j+k_3} - \left(\mathbf{1}_d'\mathbf{X}_j\right)\mathbf{X}_{j+k_2}'\mathbf{X}_{j+k_3} + 2\mathbf{1}_d'\left(\mathbf{X}_j.\mathbf{X}_{j+k_2}.\mathbf{X}_{j+k_3}\right), \end{aligned} \tag{12}$$

where $\mathbf{A}.\mathbf{B}$ means the element by element product of the vectors $\mathbf{A}$ and $\mathbf{B}$ and where $j+k_m$ is taken as $j+k_m \pmod q$ when using pitch class vectors. Equation (12) requires $O(d)$ operations for each element of $\mathbf{X}_{\mathrm{e}}^{(3)}$.

The *relative triad expectation matrix* $\hat{\mathbf{X}}_{\mathrm{e}}^{(2)}$ indicates the expected number of ordered tone triples that will be heard as corresponding to any given triad of relative pitches. It is useful for comparing the triadic structures of two or more pitch collections, regardless of transposition. For example, to compare the number of triad types two pitch collections have in common; or to compare pitch collections by the number, and tuning accuracy, of a specific set of privileged triads they each contain (for a specific application, see Example 6.4, which compares thousands of scale tunings against a just intonation triad).

Summing $\mathbf{X}_{\mathrm{e}}^{(3)}$ over j gives the relative triad expectation matrix $\hat{\mathbf{X}}_{\mathrm{e}}^{(2)} \in \mathbb{R}^{q\times q}$ indexed by $0 \le k_2, k_3 \le q-1$, with elements

$$\hat{x}_{\mathrm{e}_{k_2,k_3}} = \sum_j x_{\mathrm{e}_{j,k_2,k_3}}. \tag{13}$$

Element indices k_2 and k_3 indicate two intervals with j (which together make a triad). Assuming independence of voice saliences, the element values are the expected number of ordered tone triples heard as corresponding to that triad of relative pitches.

For example, given the pitch class vector for a four-voice minor triad with a doubled third $(0, 3, 3, 7)$ and a weighting vector of $(1, 1, 1, 1)$, the resulting relative triad expectation matrix is $\hat{\mathbf{X}}_{\mathrm{e}}^{(2)} = \begin{pmatrix} 0&0&0&0&2&0&0&0&0&2&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&2&0&0&0&2&0&0&0&0\\ 2&0&0&0&0&0&0&0&0&2&0&0\\ 0&0&0&0&0&0&0&0&2&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&2&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&2&0&0&2&0&0&0\\ 2&0&0&0&2&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0 \end{pmatrix}$. This example is indexed from top to bottom by $k_2 = 0, 1, \ldots, 11$, and from left to right by $k_3 = 0, 1, \ldots, 11$. The first row shows there are two ordered tone triples with the triadic structure $\{j, j+0, j+4\}$ (tone triples $(2, 3, 4)$, and $(3, 2, 4)$); and two ordered tone triples with the triadic structure

$\{j, j+0, j+7\}$ (triples $(2,3,1)$ and $(3,2,1)$). Row 4 shows there are two ordered tone triples containing the triadic structure $\{j, j+3, j+3\}$ (triples $(1,2,3)$ and $(1,3,2)$); and two ordered tone triples with the triadic structure $\{j, j+3, j+7\}$ (triples $(1,2,4)$ and $(1,3,4)$). And so forth.

4.4. r-ad expectation arrays

The definitions and techniques of Sections 4.1–4.3 can be generalised to an array with any number of dimensions. An *absolute r-ad expectation array*, $\mathbf{X}_{\mathrm{e}}^{(r)} \in \mathbb{R}^{\overbrace{q \times q \times \cdots \times q}^{r}}$, contains elements

$$x_{\mathrm{e}_{j,k_2,\ldots,k_r}} = \sum_{\substack{(i_1,\ldots,i_r)\in D^r:\\ i_n \neq i_o}} \prod_{m=1}^{r} x_{\mathrm{pcr}_{i_m, j+k_m}} \tag{14}$$

where $D = \{1, 2, \ldots, d\}$, and $k_1 = 0$. Element indices j, k_2, $\ldots$, k_r indicate the pitches j, $j+k_2$, $\ldots$, $j+k_r$; assuming the independence of voice saliences, element value indicates the expected number of ordered r-tuples of tones heard as having those r pitches.

Equation (14) is symbolically concise, but cumbersome to calculate since each element of $\mathbf{X}_{\mathrm{e}}^{(r)}$ requires $\frac{d!(r-1)}{(d-r)!}$ operations. Fortunately, the computational complexity can be reduced by algebraic manipulation as in Appendix B and by exploiting the sparsity of the arrays to calculate only non-zero values. Furthermore, due to their construction, the arrays are invariant with respect to any transposition of their k indices so only non-duplicated elements need to be calculated. To minimise memory requirements, the arrays can be stored in a sparse format.

The absolute r-ad expectation arrays can be made invariant with respect to transposition by summing over j. This creates an $(r-1)$-dimensional *relative r-ad expectation array*, $\hat{\mathbf{X}}_{\mathrm{e}}^{(r-1)} = \sum_j x_{\mathrm{e}_{j,k_2,\ldots,k_r}} \in \mathbb{R}^{\overbrace{q \times q \times \cdots \times q}^{r-1}}$ containing elements $\hat{x}_{\mathrm{e}_{k_2,\ldots,k_r}}$. Element indices $k_2, \ldots, k_r$ indicate a set of intervals with j (which together make an r-ad); assuming the independence of voice saliences, element value indicates the expected number of ordered r-tuples of tones that are heard as corresponding to that r-ad of relative pitches.

5. Metrics

The distance between a pair of vectors or arrays can be calculated with any standard metric. This section details two particular metrics (the ℓ_p and the cosine) which are used in the applications of Section 6.

It is reasonable to model the perceived pitch distance between any two tones with their absolute pitch difference (e.g., the pitch distance between tones with pitch values of 64 and 60 semitones is 4 semitones). The ℓ_p-metrics are calculated from absolute differences so they provide a natural choice for calculating the overall distance between pairs of category domain pitch vectors. When there are d different tones in each vector,

there are d different different pitch differences; the value of p determines how these are totalled (e.g., $p = 1$ gives the taxicab measure which simply adds the distances moved by the different voices; $p = 2$ gives the Euclidean measure; $p = \infty$ gives the largest distance moved by any voice). As discussed in Section 2, the use of such metrics is a well-established procedure [1, 9].

The metrics may be based on the intervals between pairs of pitch vectors in $\mathbb{R}^d$

$$\mathbf{d}_{\boldsymbol{w}}(\boldsymbol{x}, \boldsymbol{y}) = \left(\sum_{i=1}^{d} w_i (x_i - y_i)^p \right)^{1/p}, \tag{15}$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ may be two pitch vectors as in (1) or two pc-vectors as in (2). The weights w_i may be sensibly chosen to be the product of the saliences $w_i = x_{\mathrm{w}_i} y_{\mathrm{w}_i}$ from (3) [2]. The metrics may also treat the unordered pitch class intervals

$$\mathbf{d}_c(\boldsymbol{x}, \boldsymbol{y}) = \mathbf{d}_{\boldsymbol{w}}((\boldsymbol{x} - \boldsymbol{y}) \mod q, (\boldsymbol{y} - \boldsymbol{x}) \mod q). \tag{16}$$

Equation (15) provides a measure of pitch height while (16) provides a measure of pitch chroma.

To calculate the distance between two multidimensional expectation arrays $\mathbf{X}_{\mathrm{e}}^{(r)}$ and $\mathbf{Y}_{\mathrm{e}}^{(r)} \in \mathbb{R}^{\overbrace{q \times q \times \cdots \times q}^{r}}$, the ℓ_p-metrics can be applied in an element by element fashion. The simplest way to write this is to reshape the r-dimensional matrices into column vectors $\boldsymbol{x}$ and $\boldsymbol{y} \in \mathbb{R}^{q^r}$ which may be applied in (15). It may also be convenient to normalise the resulting distance to the interval $[0, 1]$, in which case every element of $\boldsymbol{x}$ can be normalised by $\frac{1}{2||\mathbf{X}_{\mathrm{e}}^{(r)}||_p}$ and every element of $\boldsymbol{y}$ can be normalised by $\frac{1}{2||\mathbf{Y}_{\mathrm{e}}^{(r)}||_p}$.

The cosine metric between two vectors $\boldsymbol{x}$ and $\boldsymbol{y} \in \mathbb{R}^d$ is

$$\mathbf{d}_{\cos}(\boldsymbol{x}, \boldsymbol{y}) = 1 - \frac{\boldsymbol{x}'\boldsymbol{y}}{\sqrt{(\boldsymbol{x}'\boldsymbol{x})(\boldsymbol{y}'\boldsymbol{y})}}, \tag{17}$$

where $'$ denotes the transpose operator. This may be applied to pitch vectors or to pc-vectors. It may also be applied to multidimensional expectation arrays in an element by element fashion by reshaping the arrays into column vectors.

Use of the cosine metric on interval vectors is an established procedure [12, 13] and, for expectation arrays, its meaning is easier to discern than that of the ℓ_p-metrics: It gives a normalised value for the expected number of ways in which each different r-ad in one pitch collection can be matched to a corresponding r-ad in another pitch collection. For example, consider the absolute triad expectation arrays for the scales C major and D major, where each tone has a salience of 1 and no probabilistic smoothing is applied. The numerator of the division counts the number of triad matches: both contain the triad {G, A, B}, which gives a count of 1; both contain the triad {A, C, E}, which increases the count to 2; both contain the triad {A, B, E}, which gives a cumulative total of 3; and so on, for all possible triads. The denominator of the division then normalises the value to the interval $[0, 1]$. Similarly, for a relative triad expectation array, both C major and D major contain three root-position major triads each, so there are a total of 9 ways they can be matched; both contain one root-position diminished triad each, so there is 1 way they can be matched, making a cumulative total of 10; and so on, for all possible relative triads. The denominator of the division again normalises.

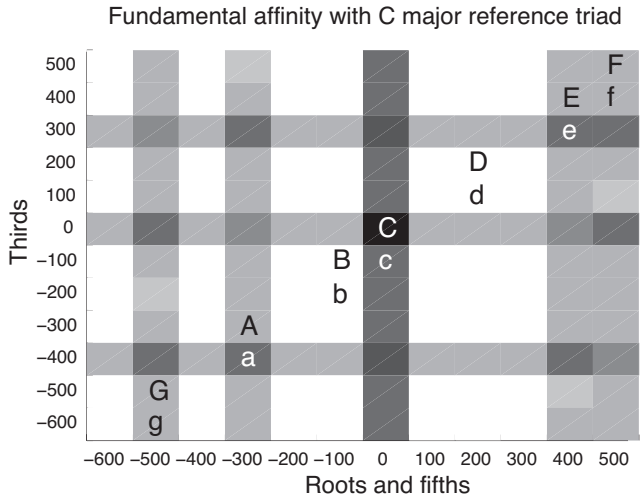

Figure 2. Fundamental pitch affinities of a C major reference triad with all possible 12-EDO triads that contain a perfect fifth. (Fundamental affinity is here modelled with a cosine metric on absolute monad expectation arrays embedding the fundamental pitches of each triad's tones; three cents standard deviation Gaussian smoothing has been used.) The horizontal axis shows the pitch distance from the reference triad's root and fifth, the vertical axis shows the pitch distance from the reference triad's third. The spatial distance between any two triads indicates their Euclidean voice-leading distance. The greyscale indicates the fundamental pitch affinity with the reference triad (the darker, the greater the fundamental affinity). Several common triads are labelled, capital letters represent major chords and small letters are minor.

6. Applications

This section provides some applications of the embeddings and metrics discussed in this paper. The MATLAB routines used to calculate them can be downloaded from `http://eceserv0.ece.wisc.edu/~sethares/pitchmetrics.html`.

6.1. Tonal distances

The pitch similarity of two chords can be modelled as a linear combination of *voice-leading distance* and *fundamental pitch affinity*: the first can be calculated by applying metrics (15) and (16) to pitch vectors; the second by applying a metric (e.g., cosine) to their absolute monad embeddings. This gives $d+4$ free parameters whose values may be determined by experimental testing—the d weights for each voice, the value of p used in the metric, and the parameters that weight the three different distance measures.

Example 6.1 Voice-leading distance and fundamental pitch affinity. This example illustrates the difference between voice-leading distance and fundamental pitch affinity. Figure 2 shows the fundamental pitch affinities (the darker the greater the fundamental affinity) between a 12-EDO reference major triad (with three voices) and all possible 12-EDO triads containing a perfect fifth. All possible root-position major and minor triads lie upon the central diagonal, some of which are labelled.

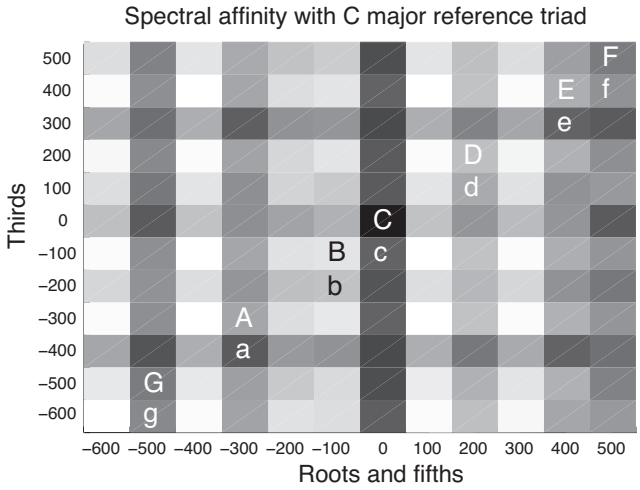

Figure 3. Spectral pitch affinities of a C major reference triad with all possible 12-EDO triads that contain a perfect fifth. (Spectral affinity is here modelled with a cosine metric on absolute monad expectation arrays embedding the first ten partials of each triad's tones; three cents standard deviation Gaussian smoothing has been used.) The horizontal axis shows the pitch distance from the reference triad's root and fifth, the vertical axis shows the pitch distance from the reference triad's third. The spatial distance between any two triads indicates their Euclidean voice-leading distance. The greyscale indicates the spectral pitch affinity with the reference triad (the darker, the greater the spectral affinity). Several common triads are labelled, capital letters represent major chords and small letters are minor.

Observe how there are local maxima of fundamental pitch affinity at those triads that have common tones with the reference C major triad (e.g., F major and A♭ major), and that the greatest maxima occur at triads that have two common tones with the reference C major triad (e.g., c minor, e minor, and a minor—which correspond to the Riemannian transformations P, L, and R). A linear combination of voice-leading distance and fundamental pitch affinity may, therefore, provide an effective measure of the overall pitch similarity of different chords [3].

Any complex tone or chord produces a large number of spectral and virtual pitch responses [5, 6], which suggests that the distances between collections of spectral or virtual pitches may provide an effective model for the perceived tonal affinity of tones or chords [2, 3]. There are so many of these pitches, it is unlikely they can be mentally categorised; for this reason, the appropriate distance function is a metric on pitch, not category, domain embeddings.

Example 6.2 Voice-leading distance and spectral pitch affinity. This example illustrates the difference between spectral pitch affinity and voice-leading distance and, comparing it with Example 6.1, the difference between the spectral and fundamental pitch affinities. Figure 3 shows the spectral pitch affinities (darker colour indicates greater spectral affinity) between a 12-EDO reference major triad (with three voices) and all possible 12-EDO triads containing a perfect fifth. All possible root-position major and minor triads lie on the central diagonal, some of which are labelled.

Observe how there is a more complex patchwork of differing affinities than in Figure 2; this model suggests that the triad pair {C major, d minor} has greater spectral affinity

than the neighbouring triad pair {C major, D major}; the triad pair {C major, F major} has greater spectral affinity than the neighbouring triad pair {C major, F♯ major}; the triad pair {C major, e minor} has greater spectral affinity than the neighbouring triad pair {C major, E major}; and so forth. These results seem indicative of the tonal function of these triad pairings: the latter pair in each case is typically heard as requiring resolution, the former pair in each case is not. This suggests that such metrics may provide effective models for the feelings of expectation and resolution induced by successions of chords in tonal-harmonic music [3].

6.2. Temperaments

The embeddings and metrics can be used to find effective *temperaments*, which are lower-dimensional tunings that provide good approximations of higher-dimensional tunings [14]. The *dimension* of a tuning is the minimum number of unique intervals (expressed in a $\log(f)$ measure like cents or semitones) that are required to generate, by linear combination, all of its intervals.

Many useful musical pitch collections are high-dimensional; for example, just intonation intervals and chords with frequency ratios 4:5:6 and 4:5:6:7 are three- and four-dimensional, respectively. But lower-dimensional tunings (principally one and two-dimensional) also have a number of musically useful features; notably, they facilitate modulation between keys, they can generate scales with simply patterned structures (equal step scales in the case of 1-D tunings, well-formed scales in the case of 2-D tunings [15]), and the tuning of all tones in the scale can be meaningfully controlled, by a musician, with a single parameter [16].

Given the structural advantages of low-dimensional generated scales, it is useful to find examples of such scales that also contain a high proportion of tone-tuples whose pitches approximate privileged higher-dimensional intervals and chords. A familiar example is the chromatic scale generated by the 100 cent semitone, which contains twelve triads (one for each scale degree) tuned reasonably close to the just intonation major triad; another familiar example is the meantone tuning of the diatonic scale (generated by a period of approximately 1200 cents and a generator of approximately 697 cents), which contains three major triads whose tuning is very close to the just intonation major triad. There are, however, numerous alternative—and less familiar—possibilities.

Given a privileged pitch class collection embedded in an expectation array, it is easy to calculate its distance from a set of n-EDOs (up to any given value of n).

Example 6.3 1-D approximations to 4:5:6 (JI major triad). The just intonation major triad contains all (and only) the common-practice harmonic consonances (i.e., the perfect fifth and fourth, and the major and minor thirds and sixths). It is, therefore, interesting to find tunings that produce simple scales containing many of these intervals. The just intonation major triad with tuning ratios of $4 : 5 : 6$ is approximated by $(0, 386.3, 700)$ cents. Figure 4 shows the cosine distance between the relative dyad expectation array embeddings of the JI major triad and all n-EDOs from $n = 2$ to 102.

Observe that the distances approach a flat line where increasing n is no longer beneficial, and that the most prominent minima fall at the familiar 12-EDO and at other alternative n-EDO's (e.g., 19-, 22-, 31-, 34-, 41-, 46-, and 53-EDO) that are well-known in the microtonal literature.

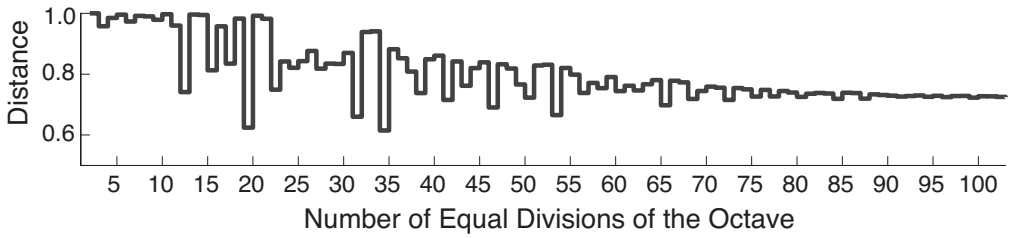

Figure 4. The distance (using the cosine metric on relative dyad expectation embeddings with a Gaussian smoothing kernel of 3 cents standard deviation) between a just intonation major triad $(0, 386.3, 702)$ and all n-edos from $n = 2$ to $n = 102$.

A two-dimensional tuning has two generating intervals with sizes, in $\log(f)$, denoted α and β. All intervals in the tuning can be generated by α and β. A *β-chain* is generated by stacking integer multiples of β for all integers in a finite range of values, so a 19-tone β-chain might consist of the notes $j\alpha-9\beta, j\alpha-8\beta, \ldots, j\alpha+8\beta, j\alpha+9\beta$. Given an arbitrary set of privileged intervals with a period of repetition ρ (typically 1200 cents), how can similar two-dimensional tunings be found? It is logical to make the tuning of $\alpha = \rho/n$, for $n \in \mathbb{N}$. For a given α, the procedure is to generate β-chains of a given cardinality and to iterate the size of the β-tuning over the desired range. At each iteration, the distance to the set of privileged intervals is measured using the relative dyad expectation embeddings and a cosine metric.

Example 6.4 2-D approximations to 4:5:6 (JI major triad). Figure 5 shows the distance between the relative dyad embeddings of a just intonation major triad and 19-tone β-tunings ranging over $0 \leq \beta \leq 1199.9$ cents in increments of 0.1 cents. On the right-hand side, the Gaussian smoothing function has a standard deviation of 3 cents; on the left, a standard deviation of 6 cents. Note that when using a single smoothing width, these charts are perfectly symmetrical about the centre line passing through 0 and 600 cents because a β-chain generated by $\beta = B$ cents is identical to that generated by $\beta = \alpha - B$ (assuming α and β are in a log value such as cents) [14].

Observe the following distance minima at different β-tunings: 503.8 cents corresponds to the familiar meantone temperament; 498.3 cents to the *helmholtz* temperament; 442.9 cents to the *sensipent* temperament; 387.8 cents to the *würschmidt* temperament; 379.9 cents to the *magic* temperament; 317.1 to the *hanson* temperament; 271.6 cents to the *orson* temperament; 176.3 cents to the *tetracot* temperament (the names for each of these temperaments has been taken from [17]). It is interesting to note that the classic meantone tunings of approximately 504 (or 696) cents are deemed closer than the helmholtz tunings of approximately 498 (or 702) cents when the smoothing has 6 cents, and vice versa when the smoothing has a 3 cent standard deviation.

Figure 6 compares the distance between between a just intonation major triad and seven-tone β-chains (with β-tunings ranging from 0 to 1199.9 cents in increments of 0.1 cents) when embedded in relative dyad and relative triad expectation arrays. The left side shows triad embeddings, the right side shows dyad embeddings.

Observe that, for low cardinality generated scales (like this seven-tone scale), only a few tunings provide tone triples that are reasonably close to the just intonation major triad: the meantone generated scale ($\beta \sim 696$ cents) contains three major triads, the magic scale ($\beta \sim 820$ cents) contains two major triads, the porcupine scale ($\beta \sim 1,037$

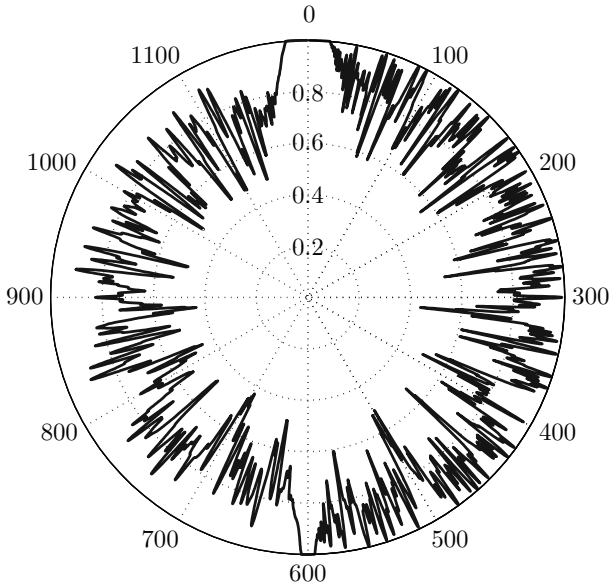

Figure 5. The cosine distance between relative dyad embeddings of a just intonation major triad $\{0, 386.3, 702\}$ and a 19-note β-chain whose β-tuning ranges from 0 to 1,199.9 cents. The smoothing is Gaussian with standard deviations of 6 cents (left side), and 3 cents (right side).

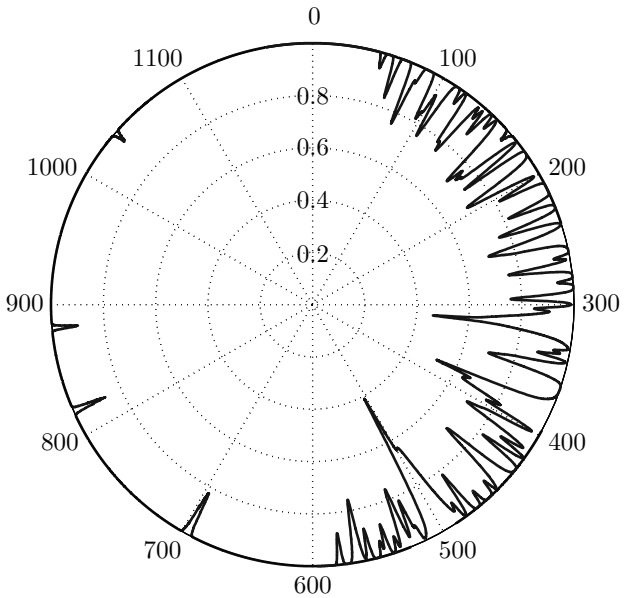

Figure 6. The cosine distance between relative dyad embeddings (right) and relative triad embeddings (left) of a just intonation major triad $\{0, 386.3, 702\}$ and a 7-tone β-chain whose β-tuning ranges from 0 to 1,199.9 cents. The smoothing is Gaussian with a standard deviation of 3 cents.

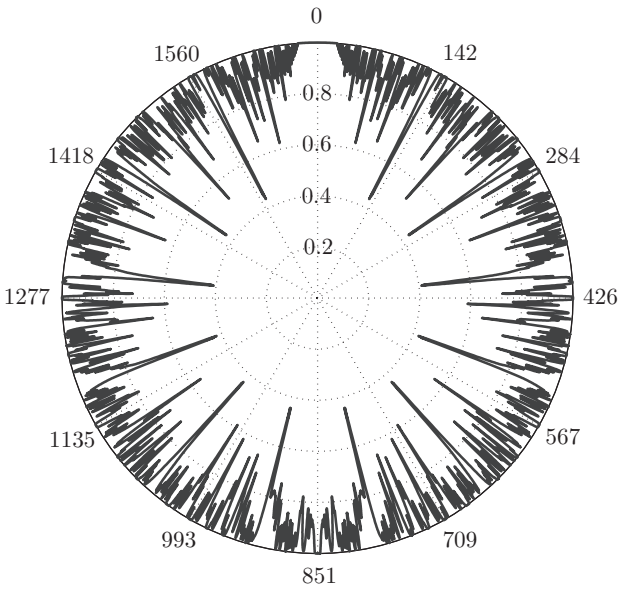

Figure 7. The cosine distance (using a Gaussian smoothing kernel with a 3 cents standard deviation) between a just intonation Bohlen-Pierce "major" triad $\{0, 884.4, 1466.9\}$, with a period of 1902 cents, and a 19-tone β-chain whose β-tuning ranges from 0 to 1901.9 cents.

cents) contains two major triads (but with less accurate tuning than the magic), the hanson scale ($\beta \sim 883$ cents) scale contains only one major triad (tuned extremely close to just intonation). As the cardinality of the β-chain is increased, the distances between the triadic embeddings approach those of the dyadic.

Example 6.5 2-D approximations to 3:5:7 (7-limit Bohlen-Pierce triad). The above two examples have used familiar tonal structures (the octave of 1200 cents and the major triad), but the methods are equally applicable to any alternative structure. One such is the Bohlen-Pierce scale, which is intended for spectra containing only odd numbered harmonics. It has a period of 3/1 (the "tritave"), which is approximated by 1902 cents. The $3 : 5 : 7$ triad, which is approximated by $\{0, 884.4, 1466.9\}$ cents, is treated as a consonance. Figure 7 shows the distance of a β-chain of 19 notes with $0 \leq \beta \leq 1901.9$ cents with a Gaussian smoothing of 3 cents standard deviation. The closest tuning is found at 439.5 cents, which is almost equivalent to $3 \times 1902/13$ and so corresponds to the 13-equal divisions of the tritave tuning suggested by Bohlen and Pierce.

6.3. Pitch set theory

There is a rich heritage of measures used to determine the distance between pitch collections in musical set theory, but these measures are typically predicated on the use of 12-tone equal temperament. Expectation arrays can be used to measure the distance between pitch collections in any tuning (up to the pitch granularity determined by q) as well as taking into account perceptual uncertainties.

The relative dyad embedding is of the $\mathrm{T}_n\mathrm{I}$ type—that is, it is invariant with respect to

Table 2. Cosine distances between a selection of pc-sets related by Z-relation, inversion, and transposition. Distances calculated with relative embeddings are in the lower triangle, absolute embeddings in the upper triangle; dyad embeddings on the top line, triad embeddings on the second line.

		$(0,1,4,6)$	$(0,1,3,7)$	$(0,2,5,6)$	$(1,2,5,7)$
$(0,1,4,6)$	dyad	0	.833	.833	1
	triad	0	1	1	1
$(0,1,3,7)$	dyad	0	0	1	.833
	triad	1	0	1	1
$(0,2,5,6)$	dyad	0	0	0	.833
	triad	1	0.5	0	1
$(1,2,5,7)$	dyad	0	0	0	0
	triad	0	1	1	0

transposition and inversion of the pitch collection it is derived from. It is also invariant over Z-relations (Z-related collections, such as $\{0,1,4,6\}$ and $\{0,1,3,7\}$, have the same interval content but are not related by transposition or inversion [10]). Relative triad (and higher-ad) embeddings are invariant only with respect to transposition—that is they are of the T_n type. The absolute embeddings have no invariances.

Example 6.6 Distances between pc-sets related by Z-relation, inversion, and transposition. Table 2 shows the cosine distances between the absolute and relative dyad and triad embeddings of pitch class vector $(0,1,4,6)$, its Z-relation $(0,1,3,7)$, its inversion $(0,1,3,7)$, and its transposition $(1,2,5,7)$. Distances calculated from absolute embeddings are in the top-right triangle, while those calculated from relative embeddings are in the bottom-left triangle. In each case, the upper number is the distance calculated using dyad embeddings, the lower number with triad embeddings.

A model of overall similarity could be calculated as a linear combination of absolute and relative embeddings of differing dimensions.

7. Discussion

This paper has presented a novel family of embeddings and metrics for determining the distance between pitch collections. The embeddings are based upon psychoacoustic principles (through the use of Gaussian smoothing) and may be useful as components in broader models of the perception and cognition of music. Indeed, to model any specific aspect of musical perception, a variety of appropriate embeddings may be linearly combined, with their weightings, the weightings of the tone saliences (if appropriate), and the type of metric, as free parameters to be determined from experimental data.

This paper has focused on expectation arrays, but the underlying pitch (class) response matrices can also be used to generate salience arrays, which give the probability of hearing any given r-ad of pitches. There may also be scope in applying Fourier transforms to the embeddings in order to determine similarities in the spectrum of n-EDOs that approximate various pitch collections.

The methods are also applicable to any domain involving the perception of discrete stimuli. An obvious example is the perception of timing in rhythms, with time replacing pitch so the smoothing represents perceptual or cognitive inaccuracies in timing; for

example, it might be possible to embed a rhythmic motif containing four events in a relative tetrad expectation matrix (in the time domain), and compare this with a selection of other similarly embedded rhythm patterns to find one with the closest match (i.e., one that contains the greatest number of patterns that are similar to the complete motif).

8. Acknowledgements

Thanks to Tuomas Eerola and Petri Toiviainen for allowing an early phase of this project to be developed as part of the Music, Mind and Technology Master's programme at the University of Jyväskylä, Finland; also to Margo Schulter for setting us an early challenge, the solution to which provided some important insights.

References

[1] D. Tymoczko, *Supporting online material for the geometry of musical chords*, Science 313, 72 (2006).
[2] R. Parncutt, *Harmony: A Psychoacoustical Approach*. Springer-Verlag, Berlin, 1989.
[3] A.J. Milne, *A psychoacoustic model of harmonic cadences: A preliminary report*, in *Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music (ESCOM 2009)*, Jyväskylä, Finland, 12–16 August 2009, pp. 328–337.
[4] M. Castrén, *RECREL: A Similarity Measure for Set-classes*, Sibelius Academy, Helsinki, 1994.
[5] E. Terhardt, G. Stoll, & M. Seewann, *Pitch of complex signals according to virtual-pitch theory: Tests, examples, and predictions*. J. Acoust. Soc. Am. 71 (1982), pp. 671–678.
[6] E. Zwicker, & H. Fastl, *Psychoacoustics: Facts and Models*. Springer, Berlin, 1999.
[7] J. G. Roederer, *The Physics and Psychophysics of Music*, Springer-Verlag, New York, 1994.
[8] W.A. Sethares, A.J. Milne, S. Tiedje, A. Prechtl, and J. Plamondon, *Spectral Tools for Dynamic Tonality and Audio Morphing*, Comput. Music J. 33 (2009), pp. 71–84.
[9] J. Chalmers, *Divisions of the Tetrachord*, Frog Peak Music, Hanover, NH, 1993.
[10] A. Forte, *The Structure of Atonal Music*, Yale University Press, New Haven and London, 1973.
[11] D.M. Green, J.A. Swets, *Signal Detection Theory and Psychophysics*, Wiley, New York, 1966.
[12] D.W. Rogers, *A Geometric Approach to Pcset Similarity*, Perspect. New Music 37, (1999), pp. 77–90.
[13] D. Scott, & E.J. Isaacson, *The Interval Angle: A Similarity Measure for Pitch-Class Sets*, Perspect. New Music 36 (1998), pp. 107–142
[14] A.J. Milne, W.A. Sethares, and J. Plamondon, *Tuning continua and keyboard layouts,* J. Math. Music 2 (2008), pp. 1–19.
[15] N. Carey, *Coherence and sameness in well-formed and pairwise well-formed scales*, J. Math. Music 1 (2006), pp. 79–98.
[16] A.J. Milne, W.A. Sethares, and J. Plamondon, *Isomorphic controllers and dynamic*

tuning: invariant fingering over a tuning continuum, Comput. Music J. 31 (2007), pp. 15–32.

[17] P. Erlich, *A middle path between just intonation and the equal temperaments, Part 1*, Xenharmonikon 18, (2006), pp. 159–199.

[18] B.C. Moore, B.R. Glasberg, & M.J. Shailer, *Frequency and intensity difference limens for harmonics with complex tones*, J. Acoust. Soc. Am. 75 (1984), pp. 500–561.

Appendix A. Standard deviation of Gaussian probability mass function

In a two-alternative forced-choice (2-AFC) experiment, the frequency difference limen (frequency DL) is normally defined as the value at which the true positive and false positive rates indicate a d' (also known as d prime) of approximately one (a true positive is when two tones with different frequencies are identified as having different pitches, a false positive is when two tones with the same frequency are identified as having different pitches). The value of d' is defined as the distance, in standard deviations, between the mean of the responses to the signal-plus-noise stimuli and the mean of the responses to the noise-alone stimuli (for the above test, a signal-plus-noise stimulus corresponds to two different frequencies; a noise-alone stimulus to two identical frequencies). This implies the internal response to a tone of pitch j is a Gaussian centred at j, with a standard deviation equivalent to the frequency DL at j.

Experimentally obtained data (e.g., [18]) typically give a frequency DL, for tones with harmonic partials, that is equivalent (over a broad range of musically useful frequencies) to a pitch DL of approximately 3 cents. Such results are obtained in laboratory conditions with simple stimuli and minimal time gaps between tones (hence comparisons are conducted from auditory sensory (echoic) memory, or short-term memory): in real music, tones and chords are presented as part of a complex and distracting stream of musical information, and there may be long gaps between the presentations of the tone collections (hence requiring long-term memory, which is less precise). For these reasons, it may be appropriate to treat 3 cents as a minimum standard deviation; larger values may provide more effective results in some models.

Appendix B. Algebraic reduction of expectation arrays

The general form of the terms that must be summed in the expectation arrays is

$$\sum_{\substack{(i_1,\ldots,i_r)\in D^r \\ i_j\neq i_k}} \prod_{m=1}^{r} x_{m,i_m}, \tag{B1}$$

where the indices i_m range over the integers between 1 and d. It is useful to think of x_m as a vector in $\mathbb{R}^d$ with individual elements of x_m indexed by i_m as x_{m,i_m} (so, for each product, each of the r different x_m vectors represents a column from the matrix $\mathbf{X}_{\text{pcr}}$, and each of the d different i_m values represents a row of that column vector). Observe that the sum is not taken over all possible indices, it excludes all cases where any of the

indices i_j and i_k are the same. It is this dependence that makes (B1) difficult to simplify. The first result considers the $r = 2$ case.

LEMMA B.1 *Let $A = (a_1, a_2, ...a_d)'$ and $B = (b_1, b_2, ...b_d)'$ be two column vectors in $\mathbb{R}^d$ and let $\mathbf{1} \in \mathbb{R}^d$ be the vector of all ones. Then*

$$\sum_{\substack{i,j \\ j \neq i}} a_i b_j = (\mathbf{1}'A)(\mathbf{1}'B) - A'B. \tag{B2}$$

Proof: Without the dependence among the indices,

$$\sum_i \sum_j a_i b_j = \sum_i a_i \sum_j b_j = (\mathbf{1}'A)(\mathbf{1}'B). \tag{B3}$$

The dependence between i and j in (B2) can be expanded as the difference

$$\sum_{\substack{i,j \\ j \neq i}} a_i b_j = \sum_i \sum_j a_i b_j - \sum_j a_j b_j. \tag{B4}$$

The final term in (B4) is the sum of the element by element product of A and B, which can be notated $A'B$. Substituting this and (B3) into (B4) gives (B2). ■

The $r = 3$ case proceeds similarly.

LEMMA B.2 *Let $A = (a_1, a_2, ...a_d)'$, $B = (b_1, b_2, ...b_d)'$, and $C = (c_1, c_2, ...c_d)'$ be three vectors in $\mathbb{R}^d$. Then*

$$\sum_{\substack{i,j,k \\ k \neq j, k \neq i, j \neq i}} a_i b_j c_k = (\mathbf{1}'A)(\mathbf{1}'B)(\mathbf{1}'C) - (\mathbf{1}'C)A'B - (\mathbf{1}'A)B'C - (\mathbf{1}'B)A'C + 2\mathbf{1}'(A.B.C) \tag{B5}$$

where . represents the element by element multiplication of vectors.

Proof: The dependencies between i, j, and k in (B5) can be expanded by first summing all the terms over all i, j, and k, and then subtracting out the terms that are disallowed in desired sum. Thus (B5) becomes

$$\sum_{\substack{i,j,k \\ k \neq j, k \neq i, j \neq i}} a_i b_j c_k = \sum_{i,j,k} a_i b_j c_k - \sum_{\substack{i,k \\ i \neq k}} a_i b_i c_k - \sum_{\substack{j,k \\ j \neq k}} a_k b_j c_k - \sum_{\substack{i,j \\ i \neq j}} a_i b_j c_j - \sum_i a_i b_i c_i. \tag{B6}$$

The first term in (B6) has no dependencies among the indices and so

$$\sum_{i,j,k} a_i b_j c_k = \sum_i a_i \sum_j b_j \sum_k c_k = (\mathbf{1}'A)(\mathbf{1}'B)(\mathbf{1}'C). \tag{B7}$$

The final term in (B6) is easily rewritten as

$$\sum_i a_i b_i c_i = \mathbf{1}'(A.B.C). \tag{B8}$$

Each of the three middle terms has the same form as in Lemma B.1, thus

$$\sum_{\substack{i,k \\ i\neq k}} (a_i b_i) c_k = (\mathbf{1}'C)A'B - \mathbf{1}'(A.B.C)$$

$$\sum_{\substack{j,k \\ j\neq k}} (a_k c_k) b_j = (\mathbf{1}'B)A'C - \mathbf{1}'(A.B.C)$$

$$\sum_{\substack{i,j \\ i\neq j}} a_i (b_j c_j) = (\mathbf{1}'A)B'C - \mathbf{1}'(A.B.C)$$

Substituting these along with (B7) and (B8) into (B5) gives the desired result. ■

The case for general r can now be constructed in an organized fashion from those for smaller r by following the logic of Lemma B.2. First, partition the dependencies among the indices by following the logic of (B6). This rewrites the r-dimensional sum with dependent indices in terms of

(1) an r-dimensional sum with independent indices
(2) a collection of $\binom{r}{2}$ $(r-1)$-dimensional sums, which are of the same form as the $(r-1)$-dimensional problem
(3) a collection of $\binom{r}{3}$ $(r-2)$-dimensional sums which are of the same form as the $(r-2)$-dimensional problem
(4) etc.

Since (i) can be rewritten easily (as in (B3) and (B7)), and since all the lower order problems have already been solved, the complete expression is obtained by adding all the terms with appropriate signs. For example, the $r = 4$ case is:

$$\begin{aligned}
\sum_{\substack{i,j,k,\ell \\ i\neq j, i\neq k, i\neq \ell \\ j\neq k, j\neq \ell, k\neq \ell}} a_i b_j c_k d_\ell &= (\mathbf{1}'A)(\mathbf{1}'B)(\mathbf{1}'C)(\mathbf{1}'D) \\
&\quad - A'B(\mathbf{1}'C)(\mathbf{1}'D) - A'C(\mathbf{1}'B)(\mathbf{1}'D) - A'D(\mathbf{1}'B)(\mathbf{1}'C) \\
&\quad - B'C(\mathbf{1}'A)(\mathbf{1}'D) - B'D(\mathbf{1}'A)(\mathbf{1}'C) - C'D(\mathbf{1}'A)(\mathbf{1}'B) \\
&\quad + 2\mathbf{1}'(A.B.C)(\mathbf{1}'D) + 2\mathbf{1}'(A.B.D)(\mathbf{1}'C) + 2\mathbf{1}'(A.C.D)(\mathbf{1}'B) + 2\mathbf{1}'(B.C.D)(\mathbf{1}'A) \\
&\quad + A'BC'D + A'CB'D + A'DB'C \\
&\quad - 6\mathbf{1}'(A.B.C.D)
\end{aligned} \tag{B9}$$

using the same notations as in Lemma B.2.