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

RESEARCH ARTICLE

Modelling the Similarity of Pitch Collections with Expectation Tensors

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

[a]*Computing Department, The 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, The 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 [2–4], and pitch and interval class distances [5]; 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 *expectation tensors* and shows how metrics between such tensors can model their perceived dissimilarity. Expectation tensors 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 tensors are invariant with respect to transposition). Examples are given to show how the metrics accord with musical intuition.

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

MCS/CCS/AMS Classification/CR Category numbers: 05A05; 05A10; 15A69; 60C05

*Corresponding author. Email: andymilne@tonalcentre.org

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

DOI: 10.1080/1745973YY

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 [5] for a survey); psychoacoustic models of chordal distance [2–4] treat tones or chords as collections of virtual and spectral pitches [6, 7] 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 called *expectation tensors* (a tensor is also known as a multi-way array), and associated metrics, that can be applied to the above areas. As discussed in sections 3 and 4, expectation tensors model the uncertainties of pitch perception by "smearing" each pitch over a range of possible values, and the width of the smearing can be related to experimentally determined frequency difference limens [8]. The tensors can embed either absolute or relative pitches (denoted *absolute* and *relative expectation tensors*, 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. Furthermore, tensors of any order (dimensionality) can be formed, allowing the embeddings to reflect the (absolute or relative) pitch, dyad, triad, and so forth, content of the pitch collection.

The distance between expectation tensors of the same order can be determined with any standard metric (such as $\mathcal{L}_p$ or cosine). A discussion of how such metrics can be applied and interpreted is found in section 5. In section 6, a number of applications of the metrics are given, and it is shown how distances between different pairs of embeddings (absolute and relative of differing orders) may be combined to produce more informative models of the similarity of pitch collections.

To avoid confusion, it is worth making some definitions explicit. A *tone* is defined as a periodic sound stimulus that may be characterised by its fundamental frequency f (or by $\log f$); a complex tone may contain many such periodic stimuli. A *pitch* is the perceptual response (auditory sensation) that is linearly related to the log frequency of a tone. A *pitch-class* is an equivalence class of all pitches that are *periods* apart—a period being a pitch difference over which pitch equivalence is perceived to exist (typically the octave). A generalised definition that extends the methods to other domains can be found in section 7.

The probability of hearing the pitch of a tone is, following Parncutt [2], denoted *salience*. Two assumptions are made to simplify the analysis: any given tone can be heard as having no more than one pitch (or pitch-class) 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.

This paper makes use of tensors and tensor notation: to aid readers unfamiliar with tensors, a brief introduction is provided in Appendix B in the online supplementary to this article. In the main text, element-level summations have also been provided to aid comprehension.

2. Category domain embeddings

Category domain embeddings—such as the familiar pitch (class) vector—contain elements whose values indicate pitches (typically in semitones). Standard metrics between two such vectors are based only on the pitch distances between elements in matching positions in the two vectors. For this reason, such pitch metrics are meaningful only when each tone in one pitch collection has a privileged relationship with a unique tone in another pitch collection; for example, when each element (index value) represents a different category such as voice (bass, tenor, alto, soprano), or scale degree, or even metrical or ordinal position in a melody. This can occur only when there are the same number of categories in each tone collection (i.e., both pitch vectors have the same dimension).

Applying metrics to category domain vectors is a well-established technique; for example, Chalmers [9] measures the distances between differently tuned tetrachords using a variety of metrics including Euclidean $\mathcal{L}_2$, taxicab $\mathcal{L}_1$, and max-value $\mathcal{L}_\infty$ (thereby treating tetrachord scale-degrees as categories), and the 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

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	$(0, 204, 386, 498, 702, 884, 1088)$	$\mathbb{R}^{7}$

$(x, 4 + x, 7 + x)$ mod 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 tones sound the pitch classes $(0, 3, 3, 7)$ and have an associated weighting vector $(.9, .6, .6, .9)$, there is probability of .9 the first tone will be heard (in ten trials, it is expected that that tone will be heard nine times); there is a probability of .6 the second tone will be heard (in ten trials, it is expected that that tone will be heard six times).

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 example 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 [10] 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." (JI is an abbreviation of just intonation, EDO is an abbreviation of equal divisions of the octave).

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 an effective metric since this would require a direct comparison of elements in $\mathbb{R}^n$ with elements in $\mathbb{R}^m$ for $n \neq m$.[1] One strategy is to identify subsets of the elements

[1]In such a case, the Hausdorff metric could be used. This metric is noteworthy because it can be used for sets with differing cardinalities. But, because the distance between any two sets is characterised by the distance between just two points in these sets, it is inadequately sensitive as a model for perceived distance. For example, the Hausdorff distances between C-E-G and D-F♯-A and between C-E-G and C-E-A are identical.

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 to 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 it is 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).

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

$$x_{\text{pcs}_{i,j}} = x_{\text{w}_i}\, \delta(j - \big[x_{\text{pc}_i}\big]), \tag{4}$$

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 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. Naively embedding into a more finely grained pitch domain (such as $q = 1200$) is problematic. For example,

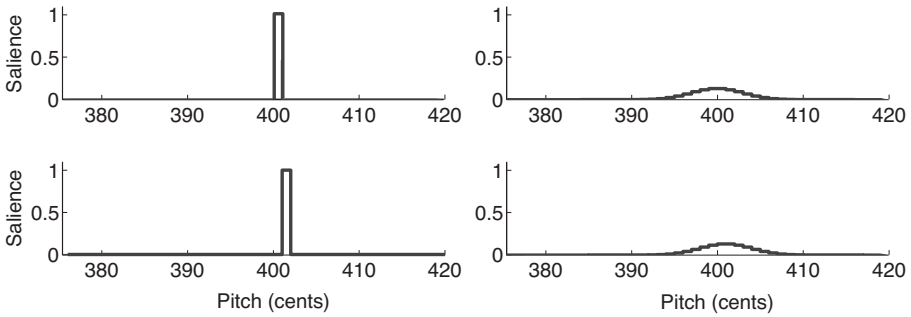

Figure 1. Pitch domain embeddings of two tones—one with a pitch of 400 cents, the other with a pitch of 401 cents. On the left, no smoothing is applied, so their distance under any standard metric is maximal; on the right, Gaussian smoothing (standard deviation of 3 cents) is applied, so their distance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

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 (almost) the same.

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 [12]. 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 in the online supplementary to this article 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* $\mathbf{X} \in \mathbb{R}^{d \times q}$ is given by

$$\boldsymbol{x}_i = \boldsymbol{x}_{\text{pcs}_i} * \boldsymbol{p} \tag{5}$$

where $\boldsymbol{x}_i$ is the ith row of $\mathbf{X}$, $\boldsymbol{x}_{\text{pcs}_i}$ is the ith row of $\mathbf{X}_{\text{pcs}}$, $\boldsymbol{p}$ is a discrete probability mass function (i.e., $p_k \geq 0$ and $\sum p_k = 1$), 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 tone (as shown in figure 1).

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

$$\mathbf{X} = \begin{pmatrix} .5 & .25 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .25 \\ 0 & 0 & .15 & .3 & .15 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & .15 & .3 & .15 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & .25 & .5 & .25 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

4. Expectation tensors

The values in the pitch class response matrix represent probabilities; this means it is possible to derive two useful types of embeddings: (a) *expectation tensors* 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 tensors* 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). (The use of an element's value to indicate the probability of hearing an interval was first suggested by Lewin in his discussion of normalised interval functions [13].) This paper focuses on expectation tensors.

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

Expectation tensors 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 and Mathematica routines have been developed to calculate the tensors discussed below; they can be downloaded from `http://eceserv0.ece.wisc.edu/~sethares/pitchmetrics.html`.

4.1. Monad expectation tensors

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 is 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 example 6.2). The elements of $\mathbf{X}_{\mathrm{e}}^{(1)}$

are derived from the elements, $x_{i,j}$, of the pitch class response matrix by

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

which is the column sum of the pitch class response matrix $\mathbf{X}$,

$$\mathbf{X}_{\mathrm{e}}^{(1)} = \mathbf{1}_d'\mathbf{X} \tag{7}$$

where $\mathbf{1}_d$ is a d-dimensional column vector of ones, and $'$ is the transpose operator. Applied to example 3.2, (7) 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 tone 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 expected 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

$$\hat{\mathbf{X}}_{\mathrm{e}}^{(0)} = \sum_{j=0}^{q-1} x_{\mathrm{e}_j} = \sum_{i=1}^{d} x_{\mathrm{w}_i} = \mathbf{1}_d'\mathbf{X}\mathbf{1}_q \tag{8}$$

where $\mathbf{1}_q$ is a q-dimensional column vector of ones. Applied to example 3.2, (8) gives $\hat{\mathbf{X}}_{\mathrm{e}}^{(0)} = 3.2$.

4.2. Dyad expectation tensors

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 (.155), the scales C major and F♯ major contain just one common dyad {B, F} and so have a large distance (.782). (These distances are calculated with a cosine metric (20) and $q = 12$.)

For dyad tensors with 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 k is given by $x_{1,j}x_{2,k}$. Similarly, the probability of hearing tone 2 as having pitch j and tone 1 as having pitch k is given by $x_{2,j}x_{1,k}$. Given two tones, the expected number of ordered tone pairs that will be heard as having pitches j and k is, therefore, given by $x_{1,j}x_{2,k} + x_{2,j}x_{1,k}$. Similarly, given three tones, there are six ordered pairs, and the

expected number of ordered tone pairs heard as having pitches j and k is given by the sum of the six probabilities.

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

$$x_{\mathrm{e}_{j,k}} = \sum_{\substack{(i_1,i_2)\in D^2:\\ i_1\neq i_2}} x_{i_1,j}\; x_{i_2,k} \tag{9}$$

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

Equation (9) requires $O(d^2)$ operations for each element. Using the tensor methods described in Appendix C in the online supplementary to this article, this can be expressed directly in terms of $\mathbf{X}$, in a way that requires only $O(d)$ operations per element,

$$\mathbf{X}_{\mathrm{e}}^{(2)} = \left(\mathbf{1}_d'\mathbf{X}\right) \otimes \left(\mathbf{1}_d'\mathbf{X}\right) - \left(\mathbf{X}'\mathbf{X}\right). \tag{10}$$

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

$$\mathbf{X}_{\mathrm{e}}^{(2)} = \begin{pmatrix} 0&0&0&2&0&0&0&1&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0\\ 2&0&0&2&0&0&0&2&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0\\ 1&0&0&2&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0 \end{pmatrix}.$$

This example is indexed from top to bottom by $j = 0, 1, \ldots, 11$, and from left to right by $k = 0, 1, \ldots, 11$. 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 pairs containing the dyad of pitches $\{3, 0\}$ (tone pairs $(2, 1)$ and $(3, 1)$); 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 $(3, 4)$). And so forth.

The *relative dyad expectation vector* $\hat{\mathbf{X}}_{\mathrm{e}}^{(1)} \in \mathbb{R}^q$ indicates the expected number of tone pairs that will be heard as corresponding to any given dyad of relative pitches (i.e., an interval). 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).

The relative dyad expectation vector is given by applying circular row shifts to $\mathbf{X}_{\mathrm{e}}^{(2)}$,

so that $k \mapsto k + j \pmod{q}$, and then summing over j, that is,

$$\hat{x}_{\mathrm{e}_k} = \sum_j x_{\mathrm{e}_{j,k+j}} \tag{11}$$

where $k + j$ is taken modulo q when pitch class vectors are used. The index k indicates an interval, of size k, with j. Assuming the independence of tone 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 unsmoothed relative dyad expectation 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 $\left\{\hat{x}_{\mathrm{e}_k} : 1 \leq k \leq \lfloor \frac{q}{2} \rfloor\right\}$ 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 6-element *interval vector* of atonal music theory [14]. The relative dyad expectation tensor can, therefore, be thought of as a generalisation of a standard interval vector that can deal meaningfully with doubled pitches and the uncertainties of pitch perception.

4.3. Triad expectation tensors

The *absolute triad expectation tensor* $\mathbf{X}_{\mathrm{e}}^{(3)}$ 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} are found in both scales) and so have a small distance (.170), 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 (20) 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, k and ℓ, respectively, are $x_{1,j}\, x_{2,k}\, x_{3,\ell}$, $x_{2,j}\, x_{1,k}\, x_{3,\ell}$, $x_{2,j}\, x_{3,k}\, x_{1,\ell}$, $x_{1,j}\, x_{3,k}\, x_{2,\ell}$, $x_{3,j}\, x_{1,k}\, x_{2,\ell}$, and $x_{3,j}\, x_{2,k}\, x_{1,\ell}$. Given three tones, the expected number of ordered tone triples heard as having pitches j, k, ℓ is given by the sum of the above probabilities.

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

$$x_{\mathrm{e}_{j,k,\ell}} = \sum_{\substack{(i_1,i_2,i_3)\in D^3:\\ i_1\neq i_2, i_1\neq i_3, i_2\neq i_3}} x_{i_1,j}\ x_{i_2,k}\ x_{i_3,\ell} \tag{12}$$

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

Equation (12) requires $O(d^3)$ operations for each element, but can be simplified to $O(d)$ by using the tensor methods of Appendix C in the online supplementary to this article:

$$\begin{aligned}\mathbf{X}_{\mathrm{e}}^{(3)} = {} & \left(\mathbf{1}_d'\mathbf{X}\right)\otimes\left(\mathbf{1}_d'\mathbf{X}\right)\otimes\left(\mathbf{1}_d'\mathbf{X}\right) - \Big(\left(\mathbf{1}_d'\mathbf{X}\right)\otimes\left(\mathbf{X}'\mathbf{X}\right)\Big)_{\langle 1,2,3\rangle} \\ & - \Big(\left(\mathbf{1}_d'\mathbf{X}\right)\otimes\left(\mathbf{X}'\mathbf{X}\right)\Big)_{\langle 2,1,3\rangle} - \Big(\left(\mathbf{1}_d'\mathbf{X}\right)\otimes\left(\mathbf{X}'\mathbf{X}\right)\Big)_{\langle 3,1,2\rangle} + 2\left(\mathbf{X}'\odot\mathbf{X}'\odot\mathbf{X}'\right)\bullet\mathbf{1}_d.\end{aligned} \tag{13}$$

Applying circular mode shifts to $\mathbf{X}_{\mathrm{e}}^{(3)}$, so that $k \mapsto k+j \pmod q$ and $\ell \mapsto \ell+j \pmod q$, and then summing over j gives the *relative triad expectation matrix* $\hat{\mathbf{X}}_{\mathrm{e}}^{(2)} \in \mathbb{R}^{q\times q}$ with elements $\hat{x}_{\mathrm{e}_{k,\ell}}$ indexed by the interval class $k, \ell \in [0, q-1]$; so

$$\hat{x}_{\mathrm{e}_{k,\ell}} = \sum_j x_{\mathrm{e}_{j,k+j,\ell+j}} \tag{14}$$

where $k+j$ and $\ell+j$ are taken modulo q when used with pitch class vectors. Element indices k and ℓ indicate two intervals, of sizes k and ℓ, with j (which together make a triad). Assuming independence of tone saliences, the element values are the expected number of ordered tone triples heard as corresponding to that triad of relative pitches.

$\hat{\mathbf{X}}_{\mathrm{e}}^{(2)}$ 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).

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 $k = 0, 1, \ldots, 11$, and from left to right by $\ell = 0, 1, \ldots, 11$. 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 tensors

The definitions and techniques of the previous sections can be generalised to a tensor of any order. An *absolute r-ad expectation tensor*, $\mathbf{X}_{\mathrm{e}}^{(r)} \in \mathbb{R}^{q^r}$, contains elements

$$x_{\mathrm{e}_{j_1,j_2,\ldots,j_r}} = \sum_{\substack{(i_1,\ldots,i_r)\in D^r:\\ i_n\neq i_p}} \prod_{m=1}^{r} x_{i_m,j_m} \tag{15}$$

where $D = \{1, 2, \ldots, d\}$. Element indices j_1, j_2, $\ldots$, j_r indicate the pitches j_1, j_2, $\ldots$, j_r; assuming the independence of tone saliences, element value indicates the expected number of ordered r-tuples of tones heard as having those r pitches. As explained in Appendix C in the online supplementary to this article, this can also be expressed directly in tensor notation:

$$\mathbf{X}_{\mathrm{e}}^{(r)} = \left(\left(\mathbf{1}_{q^r} \otimes \mathbf{E}_{d^r} \right) \circ \mathbf{X}^{\otimes r}_{\langle r+1,1,r+2,2,\ldots,\ldots,r+r,r\rangle} \right) \overset{r}{\bullet} \mathbf{1}_{d^r}. \tag{16}$$

Equations (15) and (16) are symbolically concise, but cumbersome to calculate since each element of $\mathbf{X}_{\mathrm{e}}^{(r)}$ requires $O(d^r)$ operations. Fortunately, this can be reduced to $O(d)$ by breaking (16) into subspaces, each of which can be simplified (this process is fully explained in in Appendix C in the online supplementary to this article). The computational complexity can be further reduced by exploiting the sparsity of the tensors to calculate only non-zero values; furthermore, due to their construction, the tensors are invariant with respect to any transposition of their indices, so only non-duplicated elements need to be calculated. To minimise memory requirements, the tensors can be stored in a sparse format.

The absolute r-ad expectation tensors can be made invariant with respect to transposition by circularly shifting modes $2, 3, \ldots, r$ of $\mathbf{X}_{\mathrm{e}}^{(r)}$ so that $j_m \mapsto j_m + j_1 \pmod q$ (where $m \in [2, r]$) and then summing over j_1. This creates an order-$(r-1)$ *relative r-ad expectation tensor* with elements

$$\hat{x}_{\mathrm{e}_{j_2,j_3,\ldots,j_r}} = \sum_{j_1} x_{\mathrm{e}_{j_1,j_2+j_1,\ldots,j_r+j_1}} \in \mathbb{R}^{q^{r-1}}. \tag{17}$$

Element indices $j_2, \ldots, j_r$ indicate a set of $r-1$ intervals with j_1 (which together make an r-ad); assuming the independence of tone 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 tensors can be calculated with any standard metric. This section details two particular metrics (the $\mathcal{L}_p$ 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 $\mathcal{L}_p$-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 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 $\mathbb{R}^d$:

$$\mathrm{d_w}(\boldsymbol{x}, \boldsymbol{y}) = \left(\sum_{i=1}^{d} w_i \left| x_i - y_i \right|^p \right)^{1/p} \tag{18}$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ may be two pitch vectors as in (1) or two pc-vectors as in (2), and 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:

$$\mathrm{d_c}(\boldsymbol{x}, \boldsymbol{y}) = \left(\sum_{i=1}^{d} w_i \min_{k \in \mathbb{Z}} \left| x_i - y_i - kq \right|^p \right)^{1/p}. \tag{19}$$

Equation (18) provides a measure of pitch height distance while (19) provides a measure of pitch class (or chroma) distance.

To calculate the distance between two expectation tensors $\mathbf{X}_{\mathrm{e}}^{(r)}$ and $\mathbf{Y}_{\mathrm{e}}^{(r)} \in \mathbb{R}^{\overbrace{q \times q \times \cdots \times q}^{r}}$, the $\mathcal{L}_p$-metrics can be applied in an entrywise fashion. The simplest way to write this is to reshape the tensors into column vectors $\boldsymbol{x}$ and $\boldsymbol{y} \in \mathbb{R}^{q^r}$ which may be applied in (18). 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

$$\mathrm{d_{cos}}(\boldsymbol{x}, \boldsymbol{y}) = 1 - \frac{\boldsymbol{x}'\boldsymbol{y}}{\sqrt{(\boldsymbol{x}'\boldsymbol{x})(\boldsymbol{y}'\boldsymbol{y})}}. \tag{20}$$

This may be applied to pitch vectors or to pc-vectors and, like the $\mathcal{L}_p$-metric, it may also be applied to the expectation tensors in an entrywise fashion by reshaping the arrays into column vectors.

The cosine distance between two vectors is equivalent to their uncentred correlation, and the use of such metrics is an established procedure in music theory and cognition

[15–17]. For expectation tensors, the meaning of the cosine distance is easier to discern (and is a more obvious choice) than that of the $\mathcal{L}_p$-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 tensors 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 tensor, 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.

The final choice of metric can be made a priori (guided by theory, as above) or post-hoc (as a free parameter chosen to fit empirical data).

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 perceived overall pitch distance of two chords can be modelled as a linear combination of *voice-leading distance* and *fundamental pitch distance*: the first can be calculated by applying metrics (18) and (19) directly to pitch vectors; the second by applying a metric, such as cosine, to their absolute monad embeddings (when using unsmoothed embeddings, this metric is closely related to Parncutt's *pitch commonality* [2]). This gives $d + 3$ 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 two different distance measures.

Example 6.1 Voice-leading distance and fundamental pitch distance. This example illustrates the difference between voice-leading distance and fundamental pitch distance. Figure 2 shows the fundamental pitch distances (the lighter the colour, the greater the distance) 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, and the spatial distance between them indicates their Euclidean voice-leading distance.

Observe how there are local minima of fundamental pitch distance at those triads that have common tones with the reference C-major triad (e.g., F-major and A♭-major), and that the greatest mimina 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 Rieman-

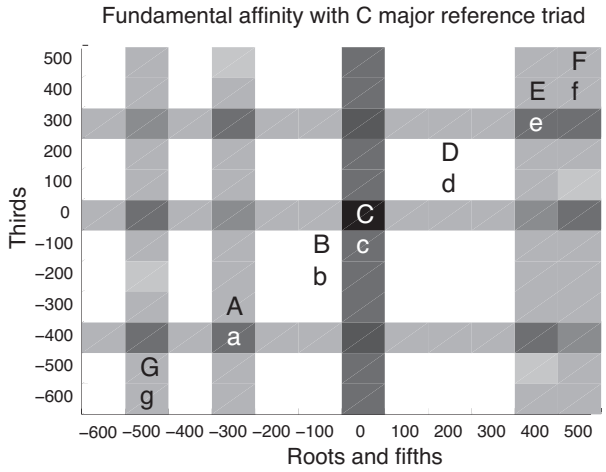

Figure 2. Fundamental pitch distances between a C-major reference triad and all possible 12-EDO triads that contain a perfect fifth. (Fundamental pitch distance is here calculated with a cosine metric on absolute monad expectation vectors embedding the fundamental pitches of each triad's tones; three 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, so the spatial distance between any two triads indicates their Euclidean voice-leading distance. The greyscale indicates the fundamental pitch distance from the reference triad (the lighter the colour, the greater the distance). Several common triads are labelled, capital letters represent major chords and small letters are minor.

nian transformations P, L, and R). A linear combination of voice-leading distance and fundamental pitch distance may, therefore, provide an effective model of the perceived overall pitch distance of different chords [3].

Any complex tone or chord produces a large number of spectral and virtual pitch responses [6, 7], which suggests that the distances between collections of spectral or virtual pitches may provide a model for the perceived affinity of tones or chords [2, 3]. There are so many of these pitches, it is unlikely they can be mentally categorised; the appropriate distance function is, therefore, a metric on pitch domain, not category domain, embeddings. Affinity is here modelled by *spectral pitch distance*: to calculate spectral pitch distance, the first ten partials of each tone in a chord are embedded in an absolute monad expectation vector, and the cosine distance between pairs of such vectors is taken; a low spectral pitch distance is hypothesized to correspond to high perceived affinity.

Example 6.2 Voice-leading distance and spectral pitch distance. This example illustrates the difference between spectral pitch distance and voice-leading distance and, comparing it with example 6.1, the difference between the spectral and fundamental pitch distances. Figure 3 shows the spectral pitch distances (lighter colour indicates greater spectral distance, and hence lower 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, and the spatial distance between them indicates their Euclidean voice-leading distance. Each

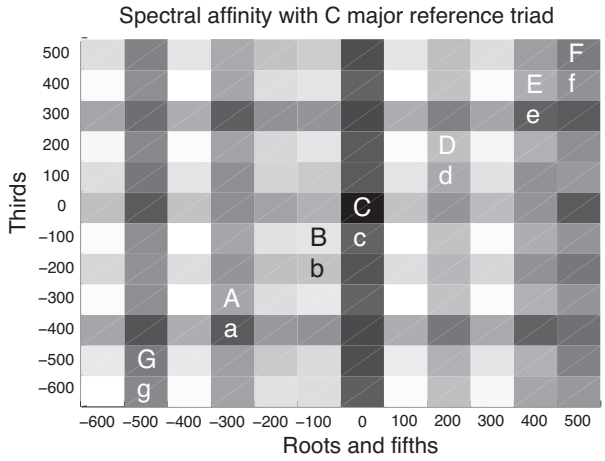

Figure 3. Affinities, as modelled by the spectral pitch distances, between a C-major reference triad and all possible 12-EDO triads that contain a perfect fifth. (Spectral pitch distance is here calculated with a cosine metric on absolute monad expectation vectors embedding the first ten partials of each triad's tones; three cents standard deviation Gaussian smoothing has been used.) The greyscale indicates the spectral pitch distance from the reference triad (the lighter the colour, the greater the distance and hence the lower the modelled affinity). In all other respects this figure is the same as figure 2.

rectangle, therefore, represents a triad pair; for example, the rectangle labelled D represents the triad pair {C-major, D-major}, and the rectangle labelled d represents the triad pair {C-major, d-minor}; the spatial distance between these two rectangles indicates the Euclidean voice-leading distance between these two triad pairs.

Observe how there is a more complex patchwork of differing distances than in figure 2; this model suggests that the triad pair {C-major, d-minor} has greater affinity than the neighbouring triad pair {C-major, D-major} (the rectangle labelled d is darker than the rectangle labelled D); the triad pair {C-major, G-major} has greater affinity than the neighbouring triad pair {C-major, G♭-major}; the triad pair {C-major, e-minor} has greater affinity than the neighbouring triad pair {C-major, E-major}; and so forth.

These patterns of differing affinities can be used to model some of the feelings of expectation and resolution induced by tonal harmony: it may be hypothesised that any pair of major or minor triads is likely to be heard as an alteration of another pair of major or minor triads that has (significantly) higher affinity and is also (significantly) close in terms of voice-leading. In figure 3, this is illustrated by two pairs of major or minor triads that have different shadings and are spatially close. For example, the triad pair {C-major, D-major} may be heard as an alteration of its higher-affinity neighbour {C-major, d-minor}—the altered tone, F♯, is resolved by continuing in the same direction as its alteration to the tone G (in a G-major or e-minor triad), thus describing a IV→V→I or IV→V→vi cadence. Similarly, {C-major, G♭-major} may be heard as an alteration of its higher-affinity neighbour {C-major, G-major}—in this case the whole triad, G♭-major, may be considered to be altered (flattened) so it is resolved, in the same direction as its alteration, to F-major, thus describing a V→♭II→I cadence. Similarly, {C-major, E-major} may be be heard as an alteration of its higher-affinity neighbour {C-major,

e-minor}—the altered tone, G♯, is resolved by continuing in the same direction as its alteration to the tone A (in an a-minor or F-major triad)—thus describing a ♭III→V→i cadence or a ♭III→V→♭VI deceptive cadence.

Many other plausible examples can be found, and a similar chart can be produced with a minor triad reference. The underlying model is explored in greater detail in [3] and [4], and is the subject of ongoing research.

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 [18]. 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 [19]), and the tuning of all tones in the scale can be meaningfully controlled, by a musician, with a single parameter [20].

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 tensor, 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 JI (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 frequency ratios of 4:5:6 is approximated by $(0, 386.3, 702)$ cents. Figure 4 shows the cosine distance between the relative dyad expectation tensor 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.

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

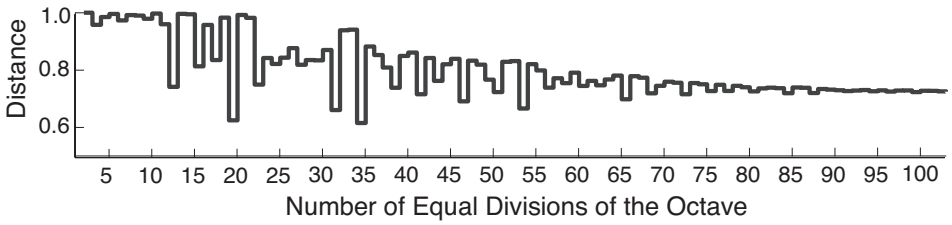

Figure 4. The distance (using the cosine metric on relative dyad expectation embeddings with a Gaussian smoothing kernel of 3 cents standard deviation) between a just intonation major triad $(0, 386.3, 702)$ and all n-edos from $n = 2$ to $n = 102$.

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 convenient to fix the tuning of α to ρ/n, for $n \in \mathbb{N}$, because this ensures the resulting generated scale repeats at the period whatever the value of β. So, once α is chosen, 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) [18].

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 [21]). 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 tensors. 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 ($\beta \sim 1,037$ cents) contains two major triads (but with less accurate tuning than the magic), the

Journal of Mathematics and Music
Vol. 00, No. 00, May 2010, 1–6

RESEARCH ARTICLE

Modelling the Similarity of Pitch Collections with Expectation Tensors
Online Supplementary: Appendices

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*)

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., [1]) 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,

*Corresponding author. Email: andymilne@tonalcentre.org

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

DOI: 10.1080/1745973YYxxxxxxxx

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. Tensors, tensor operations, and their notation

A *tensor* is a generalisation of a vector or matrix into higher *orders*. An order-0 tensor is a scalar, an order-1 tensor is a vector, an order-2 tensor is a matrix, an order-3 tensor may be thought of as a 3-dimensional array of numbers, and so forth. The *size* of a tensor of order-r $\mathbf{X} \in \mathbb{R}^{i \times j \times \cdots \times m}$ may be shown as $\overbrace{i \times j \times \cdots \times m}^{r}$, which means the first *mode* is of dimension i (it contains i entries); the second mode is of dimension j, and so forth. It is often convenient to specify the order of a tensor by its subscript so that $\mathbf{X}_{q^3}$ represents an order-3 tensor in $\mathbb{R}^{q^3}$ (which is $\mathbb{R}^{q \times q \times q}$). Lowercase italic letters such as $x_{i,j,k}$ denote scalar tensor entries, and the subscripts specify the locations. A specific permutation of a tensor's modes is indicated with a subscript in angle brackets, so if $\mathbf{X}$ is a tensor of size $i \times j \times k \times \ell$, $\mathbf{X}_{\langle 3,1,4,2 \rangle}$ has size $j \times \ell \times i \times k$.

The symbol $\circ$ denotes the *Hadamard (entrywise) product* of two tensors. If $\mathbf{C} = \mathbf{A} \circ \mathbf{B}$, then $c_{i,j,\ldots} = a_{i,j,\ldots}\, b_{i,j,\ldots}$ ($\mathbf{A}$ and $\mathbf{B}$ must be of the same size). For example,

$$\begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} \circ \begin{pmatrix} 5 & 7 \\ 6 & 8 \end{pmatrix} = \begin{pmatrix} 1{\cdot}5 & 3{\cdot}7 \\ 2{\cdot}6 & 4{\cdot}8 \end{pmatrix} = \begin{pmatrix} 5 & 21 \\ 12 & 32 \end{pmatrix}. \tag{B1}$$

The *outer product* $\otimes$ of a tensor $\mathbf{A}$ of size $i \times j$ and a tensor $\mathbf{B}$ of size $\ell \times m$ produces a tensor of size $i \times j \times \ell \times m$ containing all possible products of their elements. If $\mathbf{C} = \mathbf{A} \otimes \mathbf{B}$, then $c_{i,j,\ldots,\ell,m,\ldots} = a_{i,j,\ldots}\, b_{\ell,m,\ldots}$. For example,

$$\begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} \otimes \begin{pmatrix} 5 & 7 \\ 6 & 8 \end{pmatrix} = \left(\begin{array}{cc|cc} 1{\cdot}5 & 1{\cdot}7 & 3{\cdot}5 & 3{\cdot}7 \\ 1{\cdot}6 & 1{\cdot}8 & 3{\cdot}6 & 3{\cdot}8 \\ \hline 2{\cdot}5 & 2{\cdot}7 & 4{\cdot}5 & 4{\cdot}7 \\ 2{\cdot}6 & 2{\cdot}8 & 4{\cdot}6 & 4{\cdot}8 \end{array}\right) = \left(\begin{array}{cc|cc} 5 & 7 & 15 & 21 \\ 6 & 8 & 18 & 24 \\ \hline 10 & 14 & 20 & 28 \\ 12 & 16 & 24 & 32 \end{array}\right). \tag{B2}$$

The 2×2 partitions help to visualise the four modes of the resulting tensor: stepping from a partition to the one below increments the index of the first mode; stepping from a partition to the one on its right increments the index of the second mode; stepping down a row, within the same partition, increments the index of the third mode; stepping rightwards by a column, within the same partition, increments the index of the fourth mode. The symbol $^{\otimes r}$ denotes the rth outer power of a tensor.

The *Khatri-Rao product* $\odot$ is the "matching columnwise" Kronecker product of matrices. The Khatri-Rao product of a matrix of size $i \times n$ and a matrix of size $j \times n$ is a matrix of size $ij \times n$ (which may be interpreted as a tensor of size $i \times j \times n$). If $\mathbf{C} = \mathbf{A} \odot \mathbf{B}$, then $c_{i,j,n} = a_{i,n}\, b_{j,n}$. This can be naturally extended to successive Khatri-Rao products of matrices: if $\mathbf{F} = \mathbf{A} \odot \mathbf{B} \odot \cdots \odot \mathbf{D}$, then $f_{i,j,\ldots,\ell,n} = a_{i,n}\, b_{j,n} \cdots d_{\ell,n}$ (the rows of the

matrices, indexed here by n, must have the same dimension).[1] For example,

$$\begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} \odot \begin{pmatrix} 5 & 7 \\ 6 & 8 \end{pmatrix} = \left(\begin{array}{cc} 1{\cdot}5 & 1{\cdot}6 \\ 3{\cdot}7 & 3{\cdot}8 \\ \hline 2{\cdot}5 & 2{\cdot}6 \\ 4{\cdot}7 & 4{\cdot}8 \end{array}\right) = \left(\begin{array}{cc} 5 & 6 \\ 21 & 24 \\ \hline 10 & 12 \\ 28 & 32 \end{array}\right), \tag{B3}$$

and

$$\begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} \odot \begin{pmatrix} 5 & 7 \\ 6 & 8 \end{pmatrix} \odot \begin{pmatrix} 9 & 11 \\ 10 & 12 \end{pmatrix} = \left(\begin{array}{cc|cc} 1{\cdot}5{\cdot}9 & 3{\cdot}7{\cdot}11 & 1{\cdot}6{\cdot}9 & 3{\cdot}8{\cdot}11 \\ 1{\cdot}5{\cdot}10 & 3{\cdot}7{\cdot}12 & 1{\cdot}6{\cdot}10 & 3{\cdot}8{\cdot}12 \\ \hline 2{\cdot}5{\cdot}9 & 4{\cdot}7{\cdot}11 & 2{\cdot}6{\cdot}9 & 4{\cdot}8{\cdot}11 \\ 2{\cdot}5{\cdot}10 & 4{\cdot}7{\cdot}12 & 2{\cdot}6{\cdot}10 & 4{\cdot}8{\cdot}12 \end{array}\right) = \left(\begin{array}{cc|cc} 45 & 231 & 54 & 264 \\ 50 & 252 & 60 & 288 \\ \hline 90 & 308 & 108 & 352 \\ 100 & 336 & 120 & 384 \end{array}\right). \tag{B4}$$

As before, the partitions indicate the resulting tensors' modes. The symbol $\odot^r$ denotes the rth Khatri-Rao power.

The inner (dot) product $\bullet$ contracts the last index of the first tensor with the first index of the second tensor: if $\mathbf{C} = \mathbf{A} \bullet \mathbf{B}$, then $c_{\ldots,i,j,\ell,m,\ldots} = \sum_k a_{\ldots,i,j,k}\ b_{k,\ell,m,\ldots}$ (the inner two modes of $\mathbf{A}$ and $\mathbf{B}$, indexed here by k, must have the same dimension). For an order-r tensor and an order-s tensor, this results in an order-$(r+s-2)$ tensor. For example,

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \bullet \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} = 1{\cdot}4 + 2{\cdot}5 + 3{\cdot}6 = 32, \tag{B5}$$

and

$$\begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} \bullet \begin{pmatrix} 5 & 7 \\ 6 & 8 \end{pmatrix} = \begin{pmatrix} 1{\cdot}5{+}3{\cdot}6 & 1{\cdot}7{+}3{\cdot}8 \\ 2{\cdot}5{+}4{\cdot}6 & 2{\cdot}7{+}4{\cdot}8 \end{pmatrix} = \begin{pmatrix} 23 & 31 \\ 34 & 46 \end{pmatrix}. \tag{B6}$$

Appendix C. Computational simplification of expectation tensors

The general form of the expectation tensors is, as shown in section 4.4,

$$x_{\mathrm{e}_{j_1,j_2,\ldots,j_r}} = \sum_{\substack{(i_1,\ldots,i_r)\in D^r:\\ i_n \neq i_p}} \prod_{m=1}^{r} x_{i_m,j_m}, \tag{C1}$$

which can be written in tensor notation as

$$\mathbf{X}_{q^r} = \left(\left(\mathbf{1}_{q^r} \otimes \mathbf{E}_{d^r} \right) \circ \mathbf{X}^{\otimes r}_{\langle r+1,1,r+2,2,\ldots,\ldots,r+r,r \rangle} \right) \overset{r}{\bullet} \mathbf{1}_{d^r} \tag{C2}$$

[1] In Mathematica, this product can be written `Outer[Times, a, b, ..., d, 1]` where the final "`1`" specifies the level at which the outer product is calculated.

where $\mathbf{1}_{q^r} \in \mathbb{R}^{q^r}$ is the tensor of all ones, the $\overset{r}{\bullet}$ inner product with $\mathbf{1}_{d^r}$ represents r successive inner products with $\mathbf{1}_d$, and E_{d^r} is constructed with elements

$$e_{i_1,i_2,\ldots,i_r} = \begin{cases} 0 & \text{if } i_n = i_p, \\ 1 & \text{otherwise.} \end{cases} \tag{C3}$$

To understand the construction in (C2), observe that the outer product $\mathbf{1}_{q^r} \otimes \mathsf{E}_{d^r}$ extends the tensor of nonrepeated indices into r additional modes, each of dimension q. Since $\mathbf{X}$ is a $d \times q$ matrix, $\mathbf{X}^{\otimes r} \in \mathbb{R}^{d\times q\times d\times q\times\cdots\times d\times q}$ is of order $2r$. The index permutation reshapes $\mathbf{X}^{\otimes r}$ into an element of $\mathbb{R}^{q^r\times d^r}$. The Hadamard product with the permuted $\mathbf{X}^{\otimes r}$, therefore, sets all entries occurring at locations with repeated indices to zero. These are precisely the entries that are excluded from the summation (C1). The rth inner product then sums over the r different d-dimensional modes to collapse to the desired tensor in $\mathbb{R}^{q^r}$.

The expression takes this form due to the constraints on which index values are summed over. Both forms (C1) and (C2) are cumbersome to calculate directly. Were there no constraint on which indices in (C1) are summed over, (C2) would take the form

$$\mathbf{X}^{\otimes r}_{\langle r+1,1,r+2,2,\ldots,\ldots,r+r,r\rangle} \overset{r}{\bullet} \mathbf{1}_{d^r}. \tag{C4}$$

This requires $(dq)^r$ multiplications, but can be reduced to q^r multiplications by rearranging it to

$$\left(\mathbf{1}_d'\mathbf{X}\right)^{\otimes r}. \tag{C5}$$

This suggests an alternative way of calculating (C2), to sum all of the terms and then subtract the terms that should be excluded.

For example, consider the $r = 2$ case. The unconstrained term is $(\mathbf{1}_d'\mathbf{X})^{\otimes 2}$ and the term corresponding to the repeated indices is $(\mathbf{X}' \odot \mathbf{X}') \bullet 1_d$, which simplifies to $\mathbf{X}'\mathbf{X}$. Hence equation (9) of the main text can be written

$$\mathbf{X}_{\mathrm{e}}^{(2)} = \left(\mathbf{1}_d'\mathbf{X}\right) \otimes \left(\mathbf{1}_d'\mathbf{X}\right) - \left(\mathbf{X}'\mathbf{X}\right). \tag{C6}$$

The process for $r = 3$ is similar. The unconstrained term is $(\mathbf{1}_d'\mathbf{X})^{\otimes 3}$. There are three terms corresponding to the $i = j$ constraint, the $j = k$ constraint and the $i = k$ constraint, each is equal to one of the transpositions of $(\mathbf{1}_d'\mathbf{X}) \otimes (\mathbf{X}'\mathbf{X})$. These have now subtracted out the $i = j = k$ constraint three times, and so $\mathbf{X}'^{\odot 3} \bullet 1_d$ must be added back in twice to compensate. Accordingly, equation (12) of the main text can be rewritten

$$\begin{aligned} \mathbf{X}_{\mathrm{e}}^{(3)} = {} & \left(\mathbf{1}_d'\mathbf{X}\right) \otimes \left(\mathbf{1}_d'\mathbf{X}\right) \otimes \left(\mathbf{1}_d'\mathbf{X}\right) - \Big(\left(\mathbf{1}_d'\mathbf{X}\right) \otimes \left(\mathbf{X}'\mathbf{X}\right)\Big)_{\langle 1,2,3\rangle} \\ & - \Big(\left(\mathbf{1}_d'\mathbf{X}\right) \otimes \left(\mathbf{X}'\mathbf{X}\right)\Big)_{\langle 2,1,3\rangle} - \Big(\left(\mathbf{1}_d'\mathbf{X}\right) \otimes \left(\mathbf{X}'\mathbf{X}\right)\Big)_{\langle 3,1,2\rangle} + 2\left(\mathbf{X}' \odot \mathbf{X}' \odot \mathbf{X}'\right) \bullet \mathbf{1}_d. \end{aligned} \tag{C7}$$

An analogous procedure can be followed for any value of r, though this becomes increasingly difficult because the number of terms grows as $r!$. Each term represents a unique minimal set of different index constraints. For example, one term $\mathcal{A}$ might have the index constraints $i_1 = i_2$ and $i_3 = i_4$. Another term $\mathcal{B}$ might have no constraint on

i_1 but have $i_2 = i_3 = i_4$. When the indices are ordered sequentially, the term can be calculated by writing each constraint as a subterm of the form

$$\mathbf{X}'^{\odot c} \bullet \mathbf{1}_d, \tag{C8}$$

where c is the number of indices in that constraint, and then taking the outer product of the different subterms. For instance, with index constraints $\mathcal{A}$, (C8) is

$$\Big((\mathbf{X}' \odot \mathbf{X}') \bullet \mathbf{1}_d\Big) \otimes \Big((\mathbf{X}' \odot \mathbf{X}') \bullet \mathbf{1}_d\Big),$$

which simplifies to

$$(\mathbf{X}'\mathbf{X}) \otimes (\mathbf{X}'\mathbf{X}).$$

With index constraints $\mathcal{B}$, (C8) is

$$(\mathbf{X}' \bullet \mathbf{1}_d) \otimes \Big((\mathbf{X}' \odot \mathbf{X}' \odot \mathbf{X}') \bullet \mathbf{1}_d\Big),$$

which simplifies to $(\mathbf{1}'_d\mathbf{X}) \otimes \big((\mathbf{X}' \odot \mathbf{X}' \odot \mathbf{X}') \bullet \mathbf{1}_d\big)$. The permutation of the indices in the constraints of a term is given by the corresponding permutation of that term's tensor. For example, the term with constraints $i_1 = i_3$ and $i_2 = i_4$ (a permutation of $\mathcal{A}$) is represented by $\big((\mathbf{X}'\mathbf{X})\otimes(\mathbf{X}'\mathbf{X})\big)_{\langle 1,3,2,4\rangle}$ while the term with constraints i_2 and $i_1 = i_3 = i_4$ (a permutation of $\mathcal{B}$) is represented by $\Big((\mathbf{1}'_d\mathbf{X}) \otimes \big((\mathbf{X}' \odot \mathbf{X}' \odot \mathbf{X}') \bullet \mathbf{1}_d\big)\Big)_{\langle 2,1,3,4\rangle}$.

For example, the $r! = 24$ terms for the $r = 4$ case can be written

$$\begin{aligned}
\mathbf{X}_{\mathrm{e}}^{(4)} = {} & (\mathbf{1}'_d\mathbf{X}) \otimes (\mathbf{1}'_d\mathbf{X}) \otimes (\mathbf{1}'_d\mathbf{X}) \otimes (\mathbf{1}'_d\mathbf{X}) \\
& - \Big((\mathbf{1}'_d\mathbf{X}) \otimes (\mathbf{1}'_d\mathbf{X}) \otimes (\mathbf{X}'\mathbf{X})\Big)_{\langle 1,2,3,4\rangle} - \Big((\mathbf{1}'_d\mathbf{X}) \otimes (\mathbf{1}'_d\mathbf{X}) \otimes (\mathbf{X}'\mathbf{X})\Big)_{\langle 1,3,2,4\rangle} \\
& - \Big((\mathbf{1}'_d\mathbf{X}) \otimes (\mathbf{1}'_d\mathbf{X}) \otimes (\mathbf{X}'\mathbf{X})\Big)_{\langle 1,4,2,3\rangle} - \Big((\mathbf{1}'_d\mathbf{X}) \otimes (\mathbf{1}'_d\mathbf{X}) \otimes (\mathbf{X}'\mathbf{X})\Big)_{\langle 2,3,1,4\rangle} \\
& - \Big((\mathbf{1}'_d\mathbf{X}) \otimes (\mathbf{1}'_d\mathbf{X}) \otimes (\mathbf{X}'\mathbf{X})\Big)_{\langle 2,4,1,3\rangle} - \Big((\mathbf{1}'_d\mathbf{X}) \otimes (\mathbf{1}'_d\mathbf{X}) \otimes (\mathbf{X}'\mathbf{X})\Big)_{\langle 3,4,1,2\rangle} \\
& + 2\Big((\mathbf{1}'_d\mathbf{X}) \otimes \Big((\mathbf{X}' \odot \mathbf{X}' \odot \mathbf{X}') \bullet \mathbf{1}_d\Big)\Big)_{\langle 1,2,3,4\rangle} + 2\Big((\mathbf{1}'_d\mathbf{X}) \otimes \Big((\mathbf{X}' \odot \mathbf{X}' \odot \mathbf{X}') \bullet \mathbf{1}_d\Big)\Big)_{\langle 2,1,3,4\rangle} \\
& + 2\Big((\mathbf{1}'_d\mathbf{X}) \otimes \Big((\mathbf{X}' \odot \mathbf{X}' \odot \mathbf{X}') \bullet \mathbf{1}_d\Big)\Big)_{\langle 3,1,2,4\rangle} + 2\Big((\mathbf{1}'_d\mathbf{X}) \otimes \Big((\mathbf{X}' \odot \mathbf{X}' \odot \mathbf{X}') \bullet \mathbf{1}_d\Big)\Big)_{\langle 4,1,2,3\rangle} \\
& + \Big((\mathbf{X}'\mathbf{X}) \otimes (\mathbf{X}'\mathbf{X})\Big)_{\langle 1,2,3,4\rangle} + \Big((\mathbf{X}'\mathbf{X}) \otimes (\mathbf{X}'\mathbf{X})\Big)_{\langle 1,3,2,4\rangle} + \Big((\mathbf{X}'\mathbf{X}) \otimes (\mathbf{X}'\mathbf{X})\Big)_{\langle 1,4,2,3\rangle} \\
& - 6\,(\mathbf{X}' \odot \mathbf{X}' \odot \mathbf{X}' \odot \mathbf{X}') \bullet \mathbf{1}_d
\end{aligned} \tag{C9}$$

While expressions like (C7) and (C9) are harder to visualise than the more compact form (C2), they can be calculated more efficiently: the unsimplified form has $O\big((dq)^r\big)$ multiplications, the simplified form has $O\big(d\,(q^r)\big)$—a ratio of $1 : d^{r-1}$. Such simplifications

are key in being able to calculate the practical examples of Section 6 of the main text, some of which use large values for d (102 in example 6.3 and 19 in example 6.5).

References

[1] 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.