

Electronic Supplementary Material S1-S4 for:

The structure of cross-cultural musical diversity

Tom Rzesutek, Patrick E. Savage, and Steven Brown

S1 – “CantoCore” Musical Classification Scheme

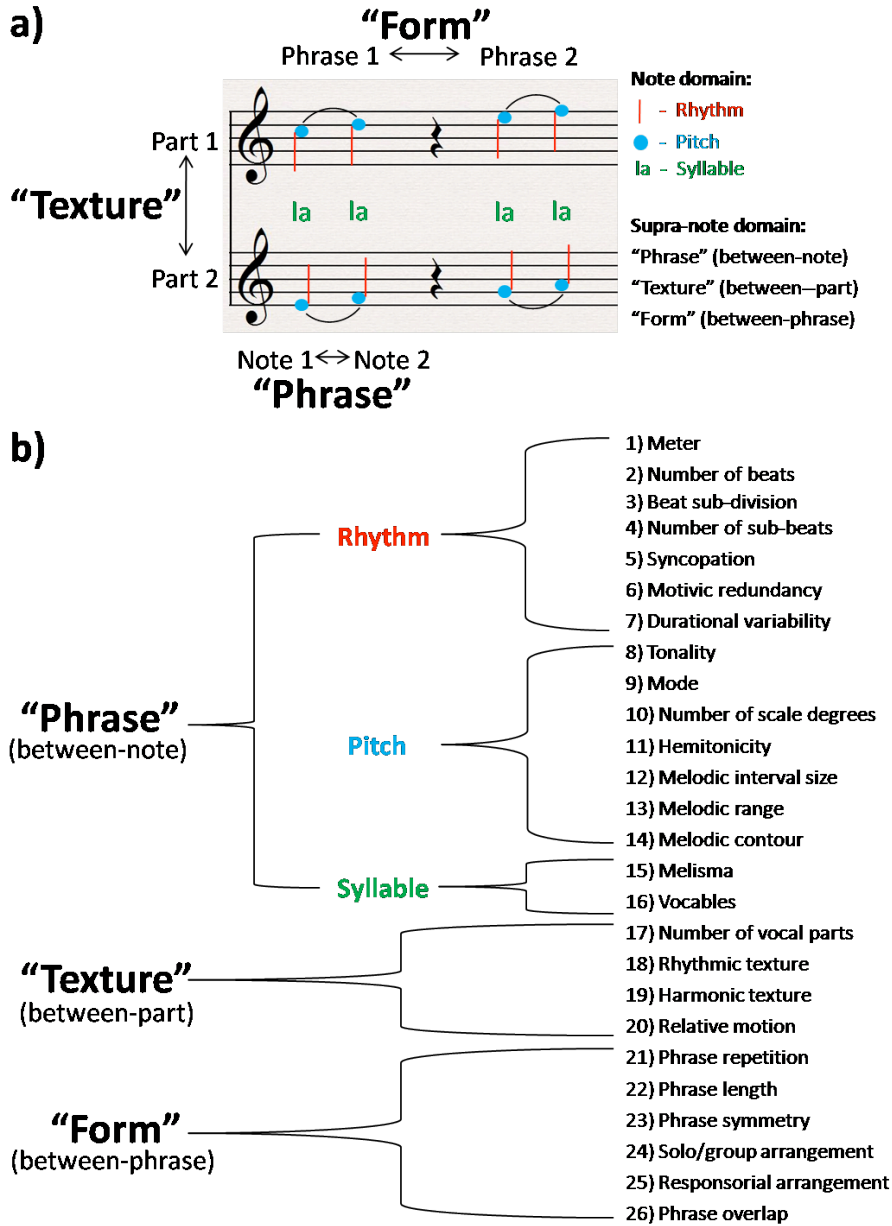


Figure S1. a) The musical hierarchy is comprised of “note” and “supra-note” domains. The three main note domains are rhythm (red), pitch (blue), and syllable (green), as represented by the sung note “la”. Interactions between notes give rise to the supra-note domains of “phrase” (the between-note level), “texture” (the between-part level) and “form” (the between-phrase level). b) The 26 structural characters that comprise the CantoCore classification scheme are organized according to these note and supra-note domains.

Character	Quantitative value	Classification
1) Meter	n/a	d) Iso-metric
2) No. of beats	n/a	b) Triple
3) Beat sub-division	n/a	c) Iso-divisive
4) No. of sub-beats	n/a	a) Simple
5) Syncopation	0%	i) Un-syncopated
6) Motivic redundancy	66% (40/61 notes derived from ♩ ♪ ♫)	iii) Highly motivic
7) Durational variability	3 unique duration values (♩, ♪, and ♫)	ii) Moderate durational variability
8) Tonality	n/a	e) Iso-tonal
9) Mode	n/a	d) Minor iso-modal
10) No. of scale degrees	6 scale degrees (A,B,C,D,E ,G)	iii) Dense scale
11) Hemitonicity	7% (4/60 intervals are the semitone between C and B)	ii) Moderately hemitonic
12) Melodic interval size	800 cents max [C-E in bars 9-10]	iii) Large intervals
13) Melodic range	1200 cents [E-E]	ii) Medium range
14) Melodic contour	n/a	cef) Descending [phrases 2, 4 & 5]; arched [phrases 6 & 8]; undulating [phrases 1,3 & 7]
15) Melisma	1 note max	i) Syllabic
16) Vocables	11% [7/61 syllables]	i) Few vocables
17) No. of vocal parts	1	i) One-part
18) Rhythmic texture	n/a	n/a) One-part
19) Harmonic texture	n/a	n/a) One-part
20) Relative motion	n/a	n/a) One-part
21) Phrase repetition	4 phrases max before a repeat	ii) Moderately repetitive
22) Phrase length	2 seconds	i) Short phrases
23) Phrase symmetry	1 (1:1 ratio of longest:shortest phrase)	i) Symmetric
24) Solo/group arrangement	n/a	c) Alternating
25) Responsorial arrangement	n/a	c) Iso-responsorial
26) Phrase overlap	0%	i) Non-overlapping

Figure S2. A sample song coding from the Shona tribe of Africa using the “CantoCore” musical coding scheme. For quantitative characters, both the raw quantitative values and categorical qualifications are shown. A transcription into musical notation is shown for illustrative purposes, but all codings are done by ear directly from recordings.

Descriptions of “CantoCore” characters

I) “MELODY” (between-note)

A) *Rhythm*

1) **METER**

Cyclic groupings of strong and weak beats into bars

- (a) *A-metric*: No consistent beat
- (b) *Hetero-metric*: There is a consistent beat, but strong and weak beats occur without a consistent pattern
- (c) *Poly-metric*: Multiple cyclic patterns of strong and weak beats coexist simultaneously (e.g., 6/8 against 3/4)
- (d) *Iso-metric*: There is a single, consistent pattern of strong and weak beats (e.g., 3/4, 6/8, 5/4, 2+2+3/8)

N.B. Songs not classified as “*iso-metric*” must be coded “?” for parameters (2-5).




2) **NUMBER OF BEATS**

The number of beats in a bar

- (i) *Duple*: Multiples of 2 (e.g., 4/4, 6/8, 2+3/8)
- (ii) *Triple*: Multiples of 3 (e.g., 3/4, 9/8, 2+2+3/8)
- (iii) *Complex*: Multiples of prime numbers greater than 3 (e.g., 5/4, 5/8, 2+2+3+2+3/8)
- (?) *A-/hetero-/poly-metric*: See (1)

3) **BEAT SUB-DIVISION**



Division of beats into sub-beat-level metric groupings

- (a) *A-divisive*: Beats are not sub-divided (e.g., a 4/4 piece containing only  and  notes)
- (b) *Hetero-divisive*: Beats are sub-divided, but the number of sub-beats per beat changes (e.g., 2+2+3/8)
- (c) *Iso-divisive*: Beats sub-divided into a consistent number of sub-beats (e.g., 6/8, a 4/4 piece containing  notes)
- (?) *A-/hetero-/poly-metric*: See (1)

N.B. Songs not classified as “*iso-divisive*” must be coded “?” for parameter (4).

4) **NUMBER OF SUB-BEATS**

The number of sub-beats in a beat

- (i) *Simple*: Multiples of 2 (e.g.,  beat divided into  note sub-beats; includes 3/4, 4/4, etc.)

- (ii) *Compound*: Multiples of 3 (e.g., ♩ beat divided into ♪ note sub-beats; includes 6/8, 9/8, “swing”, etc.)
- (iii) *Complex*: Multiples of prime numbers greater than 3 (e.g., ♩ beat divided into 5 sub-beats)
- (?) *A-/hetero-/poly-metric or a-/hetero-/poly-divisive*: See (1/3)

N.B. Songs not classified as “*iso-divisive*” must be coded “?” for parameter (4).

5) SYNCOPATION

The percentage of notes that are accented but in a metrically weak position

- (i) *Un-syncopated*: <5%
- (ii) *Moderately syncopated*: 5-20%
- (iii) *Highly syncopated*: >20%
- (?) *A-/hetero-/poly-metric*: See (1)

6) MOTIVIC REDUNDANCY

The percentage of all notes that are constructed from a single recurring rhythmic pattern

- (i) *Non-motivic*: <20%
- (ii) *Moderately motivic*: 20-50%
- (iii) *Highly motivic*: >50%

7) DURATIONAL VARIABILITY

Maximum number of different types of duration values in a song

- (i) *Low durational variability*: <3 duration values (e.g., only ♪ and ♩)
- (ii) *Moderate durational variability*: 3-4 duration values (e.g., ♪, ♩ and ♪)
- (iii) *High durational variability*: >4 duration values (e.g., ♪, ♩, ♪, ♩ and ♪)

B) Pitch

8) TONALITY

Organization of discrete pitches around one or more tonic notes

- (a) *Indeterminate a-tonal*: No discrete pitches (e.g., exclamations, heightened speech)
- (b) *Discrete a-tonal*: Discrete pitches, but no tonic
- (c) *Hetero-tonal*: Tonic modulates/shifts between phrases
- (d) *Poly-tonal*: Multiple, simultaneous tonics in different vocal parts
- (e) *Iso-tonal*: Single tonic throughout

N.B. Songs not classified as “*iso-tonal*” must be coded “?” for parameters (9-10).

9) MODE

Presence of scale degrees at a minor 3rd (250-350 cents) or major 3rd (350-450 cents) above the tonic

- (a) *A-modal*: No 3rd present
- (b) *Hetero-modal*: Both major and minor 3rd appear in separate phrases
- (c) *Poly-modal*: Both major and minor 3rd appear in the same phrase
- (d) *Minor iso-modal*: Minor 3rd only
- (e) *Major iso-modal*: Major 3rd only
- (?) *A-/hetero-/poly-tonal*: See (8)

10) NUMBER OF SCALE DEGREES

Number of scale degrees found in the scale

- (i) *Sparse scale*: <4 scale degrees
- (ii) *Moderately dense scale*: 4-5 scale degrees
- (iii) *Dense scale*: >5 scale degrees
- (?) *A-/hetero-/poly-tonal*: See (8)

11) HEMITONICITY

Percentage of melodic intervals that are semitones (50-150 cent intervals)

- (i) *Anhemitonic*: <5%
- (ii) *Moderately hemitonic*: 5-20%
- (iii) *Highly hemitonic*: >20%

12) MELODIC INTERVAL SIZE

Maximum pitch distance between successive notes within any vocal part

- (i) *Small intervals*: <350 cents (i.e., minor 3rd or less; formerly divided into “monotone”, “narrow”, and “diatonic” intervals)
- (ii) *Medium intervals*: 350-750 cents (i.e., major 3rd - perfect 5th; formerly divided into “wide” and “very wide” intervals)
- (iii) *Large intervals*: >750 cents (i.e., minor 6th or greater)

13) MELODIC RANGE

Maximum pitch distance between the highest and lowest notes within any vocal part

- (i) *Small range*: <750 cents (i.e., perfect 5th or less)
- (ii) *Medium range*: 750-1250 cents (i.e., perfect 5th - octave)
- (iii) *Large range*: >1250 cents (i.e., more than an octave)

14) MELODIC CONTOUR

Shape resulting from all changes in interval direction within a vocal part

- (a) *Horizontal*: No ascending or descending intervals
- (b) *Ascending*: Ascending intervals only
- (c) *Descending*: Descending intervals only
- (d) *U-shaped*: First descending, then ascending intervals
- (e) *Arched*: First ascending, then descending intervals
- (f) *Undulating*: Multiple changes of interval direction

C) Syllable

15) MELISMA

Maximum number of consecutive notes without articulating a new syllable

- (i) *Syllabic*: 1-2 notes
- (ii) *Mildly melismatic*: 3-5 notes
- (iii) *Strongly melismatic*: >5 notes

16) VOCABLES

The percentage of syllables containing only vowels or semi-vowels (e.g., “y”, “h”, “w”)

- (i) *Few vocables*: <20%
- (ii) *Some vocables*: 20-50%
- (iii) *Many vocables*: >50%

II “TEXTURE” (between-part)

17) NUMBER OF VOCAL PARTS

Maximum number of simultaneous vocal parts

- (i) *One-part*: 1 (formerly divided into “solo” and “unison”)
- (ii) *Two-part*: 2
- (iii) *Many-part*: >2

N.B. Songs classified as “one-part” (including both unison and solo songs) must be coded “?” for parameters **(18-20)**.

18) RHYTHMIC TEXTURE (Cantometrics Line 12)

Temporal asynchrony in the relative onsets of different vocal parts (in seconds)

- (a) *Hetero-rhythmic (heterophonic)*: 0.1–1s
- (b) *Poly-rhythmic (polyphonic)*: >1s
- (c) *Iso-rhythmic (homophonic)*: <0.1s
- (?) *One-part (monophonic)*: See **(17)**

N.B. Songs not classified as “iso-rhythmic” must be coded “?” for parameter **(19)**.

19) HARMONIC TEXTURE

Minimum harmonic interval (octave-equalized) between simultaneous vocal parts that is sustained for at least 1 second

- (i) *Dissonant*: 50-250 cents (e.g., 2^{nds}/7^{ths})
- (ii) *Consonant*: >250 cents (e.g., 3^{rds}-6^{ths})
- (?) *One-part, or poly-/hetero-rhythmic*: See **(17/18)**

20) RELATIVE MOTION

Relationship of the melodic contours (see **13**) of two simultaneous parts

- (a) *Hetero-contour (drone)*: One part is horizontal, the other changes direction

- (b) *Poly-contour (independent motion)*: Both parts have different, non-horizontal contours
- (c) *Iso-contour (parallel motion)*: Both parts have the same contour
- (?) *One-part*: See (17)

III) “FORM” (between-phrase)

21) PHRASE REPETITION

Maximum number of successive phrases before a phrase is repeated

- (i) *Non-repetitive*: >8 phrases, or no repeat at all
- (ii) *Moderately repetitive*: 3-8 phrases
- (iii) *Repetitive*: 1-2 phrases

22) PHRASE LENGTH

Maximum phrase length, in seconds

- (i) *Short phrases*: <5 s
- (ii) *Medium-length phrases*: 5-9 s
- (iii) *Long phrases*: >9 s

23) PHRASE SYMMETRY

Ratio of the length of the longest phrase in a song relative to the shortest phrase

- (i) *Symmetric*: <1.5 times the length of the shortest phrase
- (ii) *Mildly asymmetric*: 1.5-2.5 times the length of the shortest phrase
- (iii) *Very asymmetric*: >2.5 times the length of the shortest phrase

24) SOLO/GROUP ARRANGEMENT

Number of singers in each phrase

- (a) *Solo*: Only solo phrases throughout
- (b) *Mixed*: Individual phrases contain both group and solo sub-sections
- (c) *Alternating*: Alternation between distinct solo and group phrases
- (d) *Group*: Only group phrases throughout

25) RESPONSORIAL ARRANGEMENT

Alternation of phrases between different vocal parts

- (a) *A-responsorial*: No alternation between parts
- (b) *Hetero-responsorial*: Irregular alternation between parts
- (c) *Iso-responsorial*: Consistent alternation between parts

N.B. Songs classified as “*a-responsorial*” must be coded “?” for parameter (26).

Comments: See comments in (24).

26) PHRASE OVERLAP

Maximum overlap between a “call” phrase and the “response” phrase that alternates with it (as the percentage of time in which the latter phrase overlaps with the former)

- (i) *Non-overlapping*: 0%

- (ii) *Mildly overlapping*: 1–25%
- (iii) *Highly overlapping*: >25%
- (?) *A-responsorial*: See **(25)**

S2 – Song-to-song distance matrix algorithm description

a) Preparation of data

The non-independent nature of some of the characters in the CantoCore scheme results in the potential for redundancy to be introduced into the codings. For example, if the character “meter” is coded as “a-metric” (i.e., no recognisable beat), then the codings for the characters “number of beats”, “sub-beat division”, and “number of sub-beats” carry the same information, and this can overemphasise the importance of the absence of a beat. In order to reduce this potential redundancy, a “?” is put in place of redundant codings and is treated as missing data. Uncoded characters are also denoted with a “?”. Although eliminating this redundancy makes the characters more independent, further study is necessary to explore the degree of correlation between the characters and its effect on the AMOVA analysis.

b) Ordinal and nominal variables

CantoCore contains both ordinal and nominal characters. The first part of the distance-matrix algorithm separates ordinal and nominal variables, which are defined a priori by the user. Ordinal characters are coded using lower-case Roman numerals (i-iii in the case of CantoCore) and are never multi-coded. Nominal characters are coded using lower-case letters (a-f in the case of CantoCore), and can be multi-coded where appropriate.

c) Ordinal characters

Ordinal characters can have a number of possible character states, and this can vary across characters. In order to keep this consistent across ordinal characters, as well as for ordinal comparisons to be equivalent to nominal comparisons, the raw codings of the ordinal character states are converted to *scaled* values from 0 to 1 such that the minimum ordinal value is coded as 0 and the maximum ordinal value is coded as 1, with intermediate states taking intermediate values (see figure S3 a-b). In CantoCore, all ordinal characters have three possible states (i, ii, or iii). As a result, i becomes 0, ii becomes 0.5, and iii becomes 1. For a character with five states (i, ii, iii, iv, and v), the converted values would become 0, 0.25, 0.50, 0.75, and 1, respectively.

The algorithm then creates a separate pairwise distance matrix for each ordinal character by taking the absolute difference between the scaled codings for each pair of songs (Figure S3 b-c):

$$| \text{Song X} - \text{Song Y} |$$

The maximum possible difference is 1 (if the codings are maximally different) and the minimum value is 0 (if the codings are identical). For pairs of songs where one or both songs lack codings for that character (“?”), the distance is listed as NA (See Figure S3 c). The end result is a separate pairwise distance matrix for each ordinal character, from 1-j.

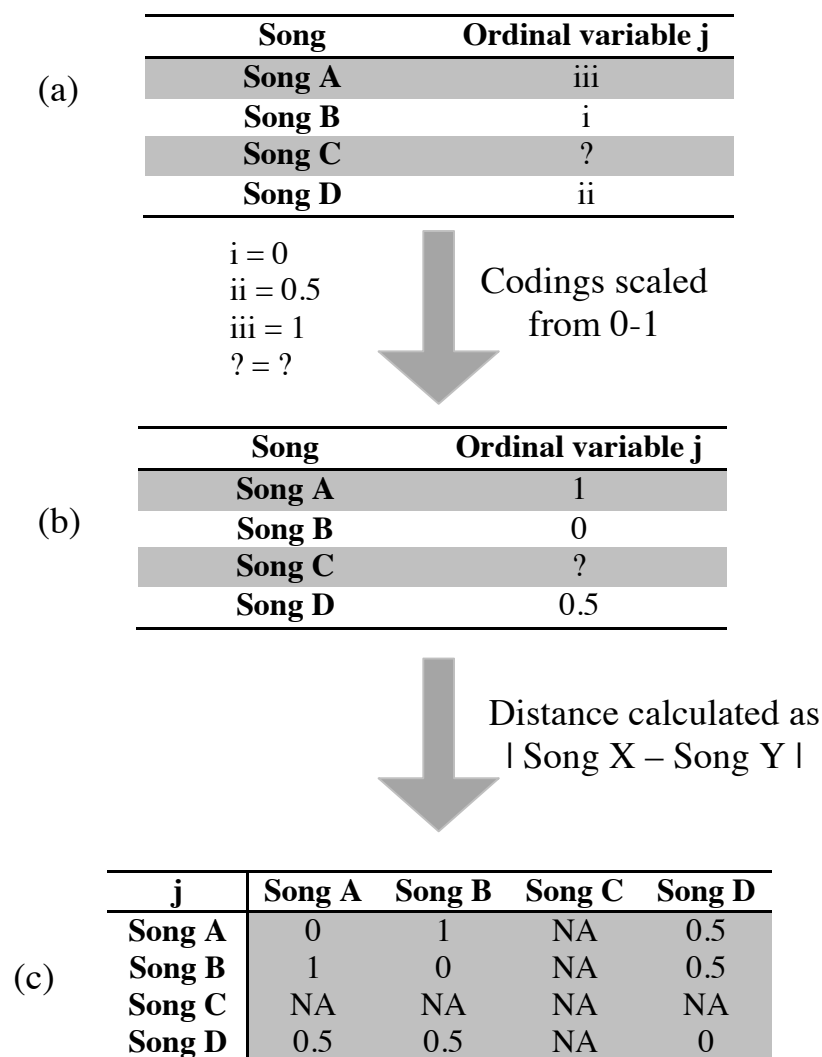


Figure S3. A sample calculation of pairwise distance across 4 songs (A-D) for a single ordinal variable (j). (a) The raw CantoCore codings. (b) The same codings after they have been converted into a scale from 0 to 1. (c) A sample distance matrix based on the absolute difference of the scaled codings for each pair of songs. Any pairwise distance involving a redundant or uncoded character (?) is denoted as NA. This overall process is repeated for each ordinal character from 1-j, resulting in j distance matrices for the song set.

d) Nominal characters

In order to accommodate the potential for multi-coded characters, the matrix of raw codings (comprised of lower-case letters) is converted into a “presence-absence” matrix, as in Busby (2006), where Y denotes the presence of a character state, and N denotes its absence (see Figure S4 a-b). Character states coded as “?” (representing redundant or missing codings) are denoted as NA when they are converted into the presence-absence matrix and are thus treated as missing data.

The conversion of scores in the presence-absence matrix into distances is based on pairwise matching of songs across all possible character states. For example, if one song contains an “a” coding (Y for character state a) and another song lacks it (N for character state a), then the program scores this as a 1, implying maximum distance between the two songs. If both songs contain a “b” coding (i.e., both are scored as Y for character state b), then the program scores this as a 0, implying minimum distance between them.

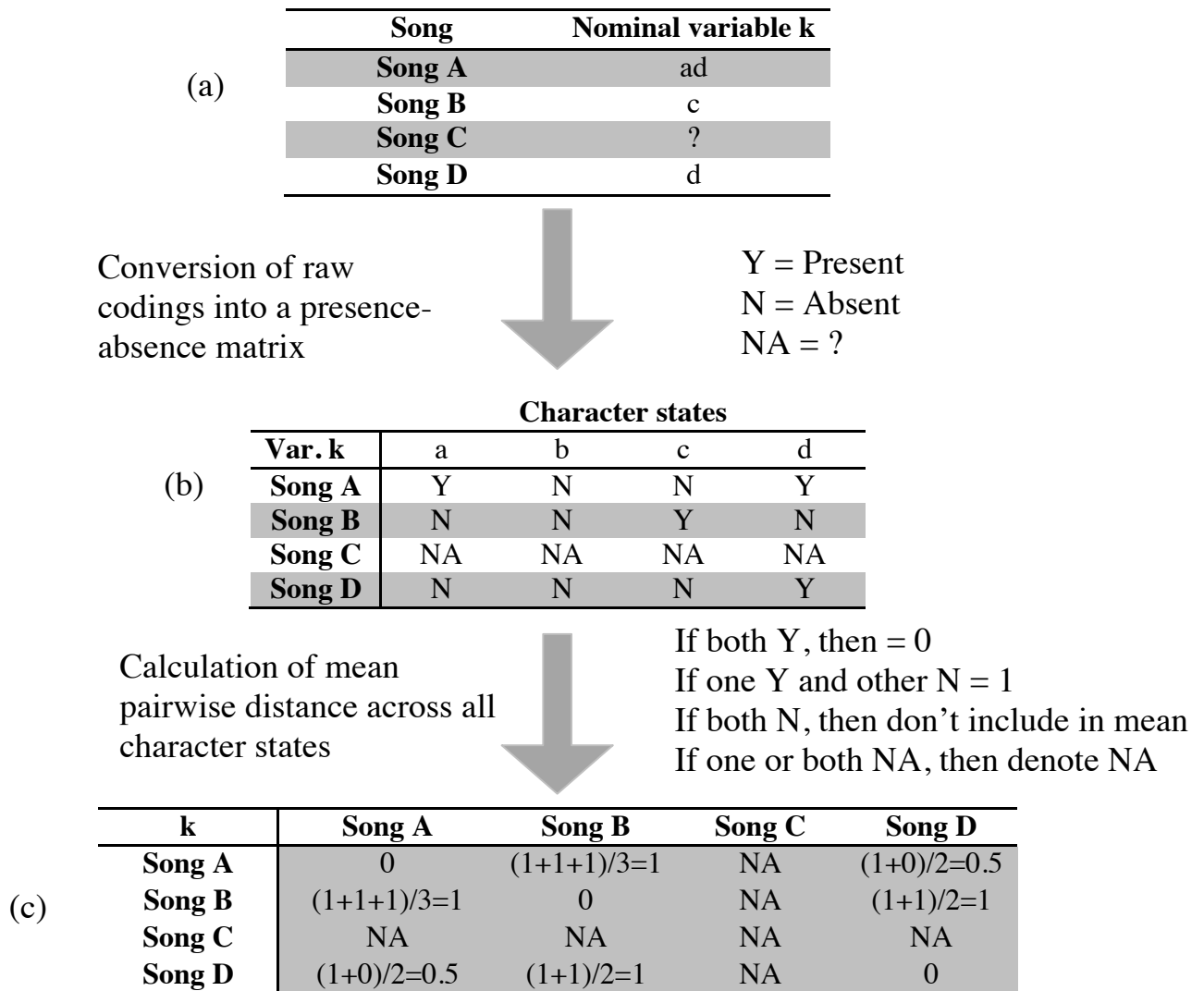


Figure S4. A sample calculation of mean distance for a nominal variable. Raw CantoCore codings (a) are first converted into a “presence-absence” matrix (b), where Y denotes the presence of a character state, N denotes its absence, and NA represents a redundant coding for each song. The mean pairwise distance is calculated by taking the pairwise distance between songs across all character states, except those involving mutual absence, which are ignored. This process is repeated for each nominal variable from 1-k.

If both songs lack a particular character state, for example *c* for Songs A and D (where both are scored as N), then this mutual absence is ignored and is not incorporated into the mean distance calculation. This is done because the mutual absence of a character state is uninformative.

Finally, to calculate the mean pairwise distance for a particular character, we take the pairwise distances between songs across all character states for that character, except those involving mutual absence, which are ignored. The occurrence of mutual absence results in some distances (means) containing fewer comparisons than others. As a result, the denominator in the mean calculation is variable. For example, the AB distance contains three comparisons while the AD distance contains only two, since the latter pair has two mutual absences compared to only one for the former pair. As for the ordinal characters, 1 is the maximum possible mean distance, and 0 is the minimum. If one or both of the songs of a pair contain NA's anywhere in their fields (because the raw coding was a "?", as with Song C), the pairwise distance is denoted as NA. A separate distance matrix is created this way for each nominal variable from 1-k.

e) Combining ordinal and nominal characters into a final distance measure

The final step of the algorithm combines information from the *j* ordinal variables with the *k* nominal variables to obtain an overall measure of distance between songs. For each pair of songs, the mean distance across all characters is taken, ignoring any distances denoted as NA. As a result, the final measure of distance incorporates information from each character equally, ignoring only redundant codings or uncoded characters.

S3 – Modal Profiles

Population	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Amis	d	a	c	a	0.5	1	0	e	d	0.5	0	0.5	1	f	0	1	0	b	?	b	1	1	0.5	c	c	0.5
Atayal	d	a	c	a	0	1	0	e	d	0.5	0	0.5	0.5	f	0	0	0	?	?	?	0.5	0.5	0	d	a	?
Ayta	a	?	?	?	?	0.5	0	e	a	0.5	0	0.5	0.5	c	0	0.5	0	?	?	?	1	0.5	0	c	c	0.5
Bunun	a	a	?	?	0	1	0.5	e	e	0.5	0	0.5	1	a	0.5	0	1	c	0.5	ca	1	1	0	b	a	?
Ibaloi	a	?	?	?	?	0.5	1	e	e	0.5	1	0.5	0.5	fe	0.5	0.5	0	?	?	?	1	1	1	c	c	0
Ifugao	a	?	?	?	?	0.5	0.5	e	?	?	0	0.5	0	f	1	0.5	1	ac	0	cb	1	0.5	0.5	d	a	?
Kankanai	d	?	?	?	?	1	0	e	e	0.5	0	0.5	0.5	f	0	0	0	?	?	?	1	0.5	0.5	d	a	?
Kavalan	d	a	c	a	0	1	0	e	d	0.5	0	0.5	0.5	f	0	0	0	?	?	?	0.5	0.5	0.5	d	a	?
Paiwan	b	?	?	?	?	1	0	e	e	0.5	0	0.5	0.5	f	0.5	0	0.5	c	0.5	a	1	1	0.5	c	a	?
Plains	a	?	?	?	?	0.5	0.5	e	d	0.5	0	0.5	1	f	0.5	0.5	0	?	?	?	1	0.5	0	d	a	?
Puyuma	d	a	c	a	0	1	0	e	d	0.5	0	0.5	1	f	0.5	1	0	?	?	?	0.5	1	0.5	d	a	?
Rukai	b	?	?	?	?	1	0.5	e	e	0.5	1	0.5	0.5	fa	0.5	0	0	?	?	?	1	1	0.5	cb	a	?
Saisiyat	a	?	?	?	?	1	0	e	d	0	0	0.5	0.5	f	0.5	0.5	0	?	?	?	1	0.5	0	c	a	?
Tao	a	?	?	?	?	0.5	1	e	d	0	0	0.5	0	a	0	0	0	?	?	?	1	1	0	d	a	?
Thao	a	?	?	?	?	1	0	e	d	0.5	0	0.5	1	f	0.5	0.5	0	?	?	?	0.5	0.5	0	d	a	?
Tsou	d	?	?	?	?	1	0	e	e	0.5	0	0.5	1	f	0.5	0.5	0.5	c	0.5	c	1	1	0.5	c	a	?

We created modal profiles by taking the most commonly coded character state for every CantoCore character in each population’s musical repertoire. Numbers across the top correspond to the 26 CantoCore categories found in S1. Nominal character states are coded as letters and multiple states are permitted. Ordinal characters are coded as numbers. “?” codings are treated as missing data, because their inclusion would carry information redundant with the coding of another character.

S4 – Neighbor-Net of modal profile distances

