Subharmonic Template Matching in Hearing and Tonal Music

Jun-ichi Takahashi

8-3-5 Nagaura-ekimae Sodegaura, Chiba, 299-0246, Japan

e-mail: takajun@joy.ocn.ne.jp

Author Note

The author and the present work have no support from anyone and no conflict of interest with any other works.

1

Abstract

In our previous paper (Takahashi, 2023), we derived a harmonic template model for pitch perception from the Sound Integration (SI) for Sound Localization in Auditory Scene Analysis, where the template was produced by multiplying and dividing prime numbers and accumulating them in frequency. In this paper, we applied the model to harmony perception. Multiplication and division by a factor of 2 gave octave equivalence, and pitch classes folded within an octave gave scales. Major and minor scales in the Pythagorean scale were the tone groups made of the accumulation or folding of the perfect fifth from the root note, respectively. It was shown that the Shruti system of Indian music was consistent with a 7-limit just intonation. Consonance and chord were considered the SI of the sounds emanating from multiple objects and defined as the reintegration of multiple musical tones into the missing fundamental. The chord progression was considered the SI of the sounds at different times and was defined as the process of destabilization and stabilization due to the motion in the perceptual potential. The driving force that produced the hierarchical structure was considered to be the nonuniformity of learning intensity in frequency-responsive neurons in template learning.

Keywords: Auditory Scene Analysis, Sound Localization, harmony, pitch perception, template matching

Introduction

Auditory perception segregates and integrates sounds through the temporal synchronization of the frequency components of the acoustic signal, localizes sound sources, and characterizes them through their spectral intensity distribution in the frequency domain. Furthermore, unlike vision, which is an energy spectrum perception in electromagnetic waves, hearing is not only a spectral perception in elastic waves but also has the special modality of mediating interaction with the behavior of others through intellectual activities such as music and language. Whereas acoustic signals are a distinct physical process, the process of their perception, hearing, remains a mystery. One of the greatest mysteries associated with hearing is music (Ball, 2010). In the human brain, there is a rewriting process from acoustic signals of a continuous physical phenomenon to a mathematically elaborated discrete structure. Let us focus on the most basic frequency structure. The brain not only perceives the acoustic spectrum as a timbre but also detects the fundamental frequency of sound from objects and perceives it as a representative value, the pitch. The pitch has the cyclic property of being grouped into similar chromas with each doubling of frequency, and in music, the pitch is further grouped into a finite structure a scale. Why are pitches concentrated in specific numbers to form scales, such as five, seven, twelve, twenty-two, or twenty-four notes? Why does a sense of harmony emerge and why do tonalities exist? And why does music induce emotions? An exhaustive amount of research has been carried out (McDermott & Hauser, 2005; Huron, 2001; Juslin & Sloboda., 2001). Although the pitch is the most fundamental concept for considering the origins of music, there remains a lively debate as to whether the mechanism of its perception is information processing in frequency space or in the time domain (Goldstein, 1973; Wightman, 1973; Terhardt, 1974; Cheveigné, 2010; Oxenham, 2013; Shamma & Dutta, 2019), even after anatomical evidence for the existence of spatially arranged auditory nerves with frequency-resolving functions (Békésy, 1949).

In our previous work, we focused on acoustic harmonic integration and derived a model of pitch perception from the Sound Localization (SL) of Auditory Scenery Analysis (ASA) (Takahashi, 2023). The most basic function of the auditory system is to separate and perceive sounds from objects in a complex acoustic environment and to provide cues for behavior (Bregman, 1994). In this process, the following processes would take place from the periphery to the center, (1) decompose the sound into its components, (2) measure the correlations among the components, (3) integrate the components, and (4) give meanings (direction, location, threat...) to each integrated signal. We modeled the Sound Integration process by the harmonic template model using a Neural Network model. We named it the

Power Series Template (PoST) model. The pitch perception is given as the success signal of SI of the sound emanating from an object mixed in the environmental sounds. The model could answer the fundamental problems of pitch perception, such as octave equivalence, missing fundamental, pitch shift, and resolved and unresolved harmonics without the help of time-theory. Acoustic signal analysis using harmonic templates not only provides for the perception of musical tones consisting of harmonics, but also for the perception of subharmonics through their inverse because the reinforcement learning is symmetrical for the fundamental and its overtones. The perception of the subharmonics is nothing other than the perception of the undertones that Riemann aspired to (Riemann, 1877). Since subharmonic components are not included in the acoustic signal from a single object, this is an overfunction in the SL of a single object. We consider that the seemingly futile subharmonic template would afford the key to the emergence of music consisting of discrete acoustic phenomena such as notes and chords. In this paper, we discuss the extension of SL in ASA by subharmonics and discuss the origin of the fundamental elements of music - scales, chords, and chord progressions - using our physiologically derived model with reference to Western music theory. Finally, the origins of why emotions are induced by music are discussed, focusing on the asymmetry of response properties in harmonics.

Model and Discussion

Power Series Templates and Pitch

We describe a brief summary of the Power Series Template (PoST) system in this section (Takahashi, 2023). Hearing began with the detection of elastic waves in viscoelastic media (including the atmosphere). To detect minute environmental acoustic signals, organisms developed dynamic amplification mechanisms. As the amplification gain increased, signal distortion, i.e. nonlinearities became non-negligible. As long as the nonlinearities are weak, perturbation theory is valid. We focus on two terms from the weaker nonlinearity and address the second- and third-order perturbation terms. Second- and third-order perturbative nonlinearities generate second and third harmonics from the reference pure tone f. In the brain, the correlation from f the fundamental component to $2f$ and $3f$ components is learned by reinforcement learning using the harmonic signals which are always generated by the nonlinear process as teacher signals. On the other hand, the correlations from the fundamental component to $\frac{1}{2}f$ and $\frac{1}{3}f$ components are also learned due to the symmetry of learning. Their chain generates power-series templates that detect synchronous correlations among $2^n f$, $3^m f$ and $2^n 3^m f$. Harmonics integrated into powers of

2 are perceived as the octave equivalence and those integrated into powers of 3 are perceived as the perfect fifth consonance.

In the inference process, each input harmonic component recalls the $2^n f$, $3^m f$ and $2^n 3^m f$ templates and is matched to each other; if an intersection is found between the 2^n and 3^m series, the value is output as a pitch. The pitch output indicates that the synchronization of the harmonic components in the input acoustic signal has been determined and that these sounds have been judged to originate from a single object (success of SI). At the same time, through the third-order nonlinearity $2f_1 - f_2$, synchronization correlations between $5f$ and $7f$ with the fundamental f are also learned using the components of the power template. As a result, a harmonic template is generated that is capable of synchronous correlation detection with the fundamental at all integers below 10, and at integers below 20 except 11, 13, 17, and 19. In the power-series template, the chain to overtones was named prime-accumulation, the chain to subharmonics (undertones) was named prime-folding and both together were named prime-power chain. We named the template matching scheme that appears in the above pitch-determination model the PoST system. In subsequent discussions of template matching, frequency is assumed to be normalized and f is omitted.

In the PoST matching scheme, we have derived, the pitch (fundamental frequency) is determined from the calculation of the input tones and does not have to be included in the input tones (Schouten, 1940; Licklider, 1954). (Missing fundamental problem) When the frequencies are shifted by the same amount in complex tones consisting of multiple harmonics, the frequencies of the templates themselves recalled by each will not match, but each matching template has a finite tolerance tuning range. The brain tries to maintain matching to a single fundamental frequency under the Gestalt principle. The matching to different fundamental frequencies at different harmonics is averaged to produce the perception of a pitch shift proportional to the frequency shift of the input sound (Schouten, 1940; Schouten, Ritsma, & Cardozo, 1962). (Pitch shift problem). The PoST system is capable of determining the pitch with two consecutive harmonics for harmonics below the 10th order. Although the template is no longer capable of determining pitch with two consecutive harmonics due to deficiencies in harmonics above the 10th order, it is possible to determine pitch with three consecutive harmonics by use of the $2f_1 - f_2$ coupling. As a result, a discontinuous jump appears in the fundamental frequency discrimination thresholds (F0DT) at the 10th. The calculation does not require the detection of a temporal structure, as the calculation is based on correlations between frequency-responsive neurons (Bernstein & Oxenham, 2003; Houtsma & Smurzynski, 1989). (Resolved and unresolved harmonics problem)

The PoST system could answer the fundamental problems of pitch perception. The harmonic template and pitch perception were shown to originate physiologically from the SL of the ASA, which occurred following the SA for the perception of small acoustic signals. At the same time, PoST allows not only matching with integer order components but also with their inverse, subharmonics. The structure of acoustic correlation with subharmonics is discussed next.

Subharmonics and Scales

Pitch perception was the SL of the sound source in ASA using harmonic groups. Subharmonics are not used for SL because the sound arriving from a single source consists of integer-order harmonics and generally does not contain subharmonics. However, when there are multiple sound sources, subharmonics matching gives affordance to the perception of acoustic correlations among sound sources.

The PoST system allows matching with the fundamental not only in 2^n3^m, but also in primes 5 and 7 through the $2f_1 - f_2$ coupling of them. The matched tones are folded back into the octave by the octave equivalence originating from the chain of PoST system, giving a pitch class. The finite set chosen from the pitch class is extended out of the octave to give the scales. The PoST system allows matching with the fundamental at all integers from 1 to 10, but only mutually prime combinations of 2, 3, 5, and 7 appear in the subharmonics because the denominator and numerator are reduced during the folding-back. The combination of frequencies produced by the folding of 2 and 3 produces the Pythagorean scale, while the folding of 2, 3, and 5 produces the just intonation. The scales using prime numbers other than 2 and 3 are called extended just intonation; it is named the prime-limit system by Partch using the maximum value of the prime number used for the folding. (Maltz, 1992; Johnston, 2010).

First, consider the subharmonic template of 2 and 3, the 3-limit system. The ratios of 2 and 3 powers that can be folded within an octave, $1<2^n3^m<2$, are listed in Table 1 in order of 3-accumulation. The table shows the combinations generated and the corresponding pitch classes, using (0,0) as the starting point C, where the numbers in the round brackets (n,m) are the order pair of the primes, here, 2 and 3. The sequence takes similar values for every 12 tones, and the classes show periodicity. The tone sequence gives the Pythagorean scale. Table 1 gives the linear representation of the circle of fifth, named the stairs of fifth.

Next, consider a 5-limit system. If we restrict the maximum degree of the 3-power chain to 2 and allow a 5-power chain, we obtain Table 1, where the pitches are represented by $2^n3^m5^l$ satisfying $1<2^n3^m5^l<2$. The combinations are arranged in order of primes and their orders. The number sequence takes similar values for each of the 12 tones and the classes show periodicity again. The sequence of tones is identical to that appearing in the 3-limit system, named the carpet of fifth, which is another representation of Tonnetz (Tymoczko, 2010).

The blue note scale used in jazz and blues uses a major chord with the addition of the third, fifth, and seventh notes lowered by a semitone. Each of these three notes is the nearest semitone of the lowest three terms 7/6, 7/5, and 7/4 of the seven-note major scale (Cutting, 2018). The blue note scale is a minimal approximation of the 7-limit system by the 5-limit system (Table 1).

The larger the order of the prime number, the finer the division of the octave. In Indian music, 22 different pitches are perceived. Twelve of these tones are systematically selected to define a group of tones called Svara, which corresponds to scale in Western music, and are used in musical performances. The Shruti system has a long history of tuning theory, but its precise musical definition is under debate (Stephens, 2017). At present, the 5-limit system and equal temperaments are widely used for tuning Shruti (Stephens, 2017; Subramanya Vasudevan, Deepak, & Ramasangu, 2022). In PoST system, the harmonic template includes seventh-order harmonics, so a just intonation system of a 7-limit system would also be natural for the auditory perception. Datta et al. conducted a cluster analysis of 150 songs performed by 53 eminent musicians and scholars covering 21 different ragas to determine the frequencies of the notes that make up the Shruti (Datta, Sengupta, & Dey, 2011). Table 2 shows the calculated values of tone clusters and the empirical assignment of the rational ratio by Datta et al. and corresponding values in various prime-limit systems. In each group, the first, second, and third columns are rational ratios, values, and the ratio to the calculated values in cent. 3- and 7-limit data are given by us and others are taken from Datta's paper. Figure 1 is the frequency distribution for the error of 22 Shrutis, where Sa is not included. Interestingly, the 3-limit system can explain the measured tone clusters with the same degree of accuracy as the 5-limit system. However, this would require more than 20 octave foldings, which would significantly complicate the rational ratio and make it impractical. At the same time, the error with the measured values is also smaller for the 5-limit system than for the 3-limit system. When analyzed with the 7-limit system, the error becomes even smaller and the error distribution becomes comparable to the best approximation by the empirical assignment of Datta et al. It would be possible that their selected performers were assumed to be aware of the tuning based on the 7-limit system. Table 3 is the combinations of prime and order reordered in their orders for 5- and 7-limit systems appearing in Table 2. There is a clear hierarchical order in the sequence of prime combinations. In both cases, the five notes of Ma Ni Ga Dha Ri appear periodically both above and below starting from Pa and Sa. In both cases, with the exception of Pa1, the correspondence of the Ma Ni Ga Dha Ri Pa groups is observed, but the index of pitch is interchanged. Pa1 is assigned to (3,-3,1,0)=40/27, adjacent to Pa2 in the 7-limit system. Another possibility is (-7,3,0,1)=189/128. We consider it inappropriate because allowing (-7,3,0,1) would not only complicate the combination but also create a jump between Ma1 (-4,1,0,1) and Pa1, which would be inconsistent with the other assignments. On the other hand, in the 5-limit system, Pa1 appears between Ri3 (1,-2,1) and Ma4 (3,-3,1), even though the value is the same as that in the 7-limit system. Pa1 is a wolf that disturbs the periodicity in Shruti.

In summary, we find that the rewritings to (A) Just intonation from the Pythagorean scale, (B) the 5-limit Shruti from the Pythagorean scale, (C) Blue Note Scale from Just intonation, (D) the 5-limit Shruti from Just intonation, and (E) the 7-limit Shruti system from the 5-limit Shruti system are the extensions of (A) from the 3-limit system to the 5-power system including the reduction of the 3-power chain, (B) from the 3-limit system to the 5-power system including the Pythagorean scale, (C) from a subclass of the 5-limit system to the 7-power system, and (D) from the small 5-limit system to the large 5-limit system, and (E) from the 5-limit system to the 7-power system including the replacement of the 3- and 5-power chains, respectively. In both cases (A)-(E), a hierarchy is observed in the combination of prime and order numbers. Does this hierarchy represent the evolution of the musical scale? We stress that our hearing had a special physiological affinity to accept the prime-limit system from the beginning. Before the development of music, 2, 3, 5, and 7 systems could already coexist in the early stages of hearing as a result of SI for SL in ASA (Takahashi, 2023). A small number of notes from all of those in the systems would have been preferred in the composition of a piece of music. As the music was played on different notes, it was perceived that subtle detuning could give different impressions, even though the pitches were almost identical. As individuals or specific groups classified and preferred them, a theory of tuning would have emerged, which fixed the combination of tones, and a scale defined by a specific number of tones would have been recognized within the group. We consider (A)-(E) to be evidence that our hearing physiologically consists of a system of 2, 3, 5, and 7, rather than representing the cultural evolution of a musical scale including the proximity to vocalization in language.

The question arises whether, if the resolution of hearing were sufficiently high, scales with even larger prime numbers than the Indian or Arab scale would be possible. Partch himself created an 11-limit tuning system (Maltz, 1992). Kubik also stated that 11/8 tones existed in African music (Kubik, 2008). However, our PoST model does not generate matching templates at prime orders above 11. We believe that the primes higher than 11 would not improve the consonance so much as $2f_1 - f_2$ coupling is needed for the integration of the 11th tone with the fundamental, the cost of which is more expensive than those for less than the 10th tone with the fundamental.

If we consider the hierarchical structure exhibited by a musical scale as a perceptual potential, its characteristics can be summarized in the following propositions.

1. the larger the prime number, the lower the stability
2. the higher the order, the lower the stability
3. both the accumulation and folding decrease the stability, but the latter does the more.
4. the effect on stability decreases in the order of the prime number and order.

As a result, the combinations are filled in the order of smaller prime numbers and smaller order. The tables show that the orders of 3 and 5 are not necessarily symmetrical with respect to 0. The stability must be nonlinear with respect to each index so the position of the apparent local minima may have shifted when hierarchically combined. The 2^n series has a high degree of freedom and the value selection is obvious for each 3^m and 5^l series. Mathematically, it defines the modulo system, musically, it gives the folding back to the octave. Both tables represent an ordinal order of stability with respect to C. Here, the contribution to stability for a unit change in prime and order is named sensory impact. We name these four laws the hierarchy principle and consider the structure of the perceptual potential and the dynamics of sound transitions on it.

Outside of the ordered structure, tones have a special structure: major and minor tonalities. According to the potential model, stability is maximal at order 0, stability decreases as the absolute value of order increases, folding has a greater decrease in stability than accumulation, and stability is periodic for 12 tones. In the Pythagorean scale given in Fig.2, the major heptatonic scale consists of seven consecutive tones from one below to above the fundamental frequency on the stair of fifth. Here, C is selected as the root note, and (-9,6) (F#), the most unstable of the seventh note created by the upward 3-accumulations (with 2-foldings) for the root note, is replaced by (2,-1) (F), which is folded back 12 notes. Also. In the minor key, (8,-5) (C#) and (10,-6) (F#), the most unstable of the seventh and eighth notes created by the downward 3-foldings (with 2-accumulations) for the root note, are replaced by (-3,2) (D), (-1,1) (G), folded back 12 tones. (A similar argument can be made for the ordinal order of tonality and sensory impact when considering 12 tones and the 5-limit system.) What is tonality is the proposition at the heart of almost all problems in Western music theory. We would like to add definitions of major scales as groups of notes reached by the 3-accumulations to the root note and minor as groups of notes reached by the 3-foldings to the root note from the standpoint of the hierarchy principle. In more general meaning, the major and minor modes in the tonality would be attributed to the perception of the overtone and undertone series.

We considered seven and twelve tones when considering tonalities. A physiological origin can account for their distribution of tone heights composed of rational ratios. Then, what is the reason for the particular preference for the numbers 5, 7, and 12 when considering scales? According to the hierarchy principle described above, in both the 3-limit and 5-limit systems, the notes are arranged in the order CGDAEBF#C#G#D#A#F. These approximately divide the octave into 12 equal intervals. Figure 3 shows the process of octave division. As the number of notes increases, each note (except D) alternately bisects the widest interval. When the number of tones is five, there are three 2- and two 3- semitone intervals respectively, and the intervals become more uniform than those for fewer tones. In this case, the scale is given the pentatonic major commonly found in world music. At the sixth note, in addition to the 2- and 3-semitone intervals, a 1-semitone interval appears, which reduces the uniformity of the interval. At seven tones, the 3-semitone intervals disappear, and uniformity is restored with five 1- and two 2-semitone intervals. Under the hierarchy principle as described above, F# may be replaced by F. The resulting scale gives a heptatonic major. At 12 tones, all intervals are 1-semitone, giving the highest uniformity. The reason Shruti has around 22 notes would be that it is the next stable point after 12 in the uniform octave division. As described above, the number of notes with a high uniformity of intervals is strongly preferred. The distribution of notes in the natural environment is random. Survival strategies would require equal attention to all pitches. Therefore, a strong preference for a particular pitch would have been avoided.

The minor scale can be discussed in the same way. The minor scales are defined as the accumulation of consecutive 3-foldings, where the most unstable notes are folded back. Due to the asymmetry of stability, the number of folded tones is two, unlike the major scale. The origins of musical scales have been discussed in terms of a preference for simple frequency ratios of sounds and their application to social communication, in particular the similarity with vocalization in languages derived from them (Gill & Purves, 2009). We propose another model of the emergence of music scales based on physiological considerations. Our model considers that template matching in Sound Integration gave rise to harmonically discrete structures in auditory perception, and that the hierarchical structure in primes and orders took on the specific numbers of tones as locally stable states for unbiased attention to the auditory environment, which is the musical scale.

Consonance and Dissonance

Whereas pitch is a result of integrating groups of tones from a single object, scales are the results of the detection of correlations among groups of tones from multiple objects.

Hearing not only judges whether the matching is good or bad, but also perceives qualitative differences in the matching results. It is generally accepted that major tones give a brighter impression and induce feelings of joy, while minor tones give a darker impression and induce feelings of sadness (Juslin & Sloboda, 2001). Similarly, when multiple tones are presented at the same time, a sense of consonance/dissonance is generated, corresponding to pleasantness or unpleasantness. The question of whether the emotional effects of major and minor tones and the impression of consonance/dissonance are innate or cultural is one of the main issues in musicology. The situation, in which numerous theories have been proposed and an active debate is still ongoing, is detailed in Harrison and Pearce's review (Harinson & Pearce, 2020).

In tonal music of Western music, the consonances between two notes are classified as perfect consonance (perfect fifth 2:3, perfect fourth 3:4), imperfect consonance (major third 4:5, minor third 5:6, major sixth 3:5, and minor sixth 5:8), and dissonance (other). These ratios fall into three groups of consecutive integers under octave equivalence: 3:4 ~ 2:3, 5:8 ~ 4:5, and 3:5 ~ 5:6. In the PoST system, two tones with consecutive two integer-ratio pitches are integrated into the missing fundamental. We consider consonance as the success of the integration of multiple sound sources under auditory correlation. From the standpoint of the subharmonic template matching view, we can formally attribute the perfect consonance to those containing one 3-power chain (2:3 group), imperfect consonance to those containing one 5-power chain accompanying with zero (4:5 group) or one (5:6 group) 3-power chain in the form of ratio, and dissonance to others (those having more than two accumulation or folding of 3 or 5) in the rational ratio of the frequencies of two tones ignoring the 2^n in the denominator and numerator.

Chord

First, we pay attention to the frequency structure of the triads (Table 4). In the C major scale, CEG(C), GBD(G), and FAC(F) are called major triads. Each has a frequency structure of 1/4*(4,5,6), 3/8*(4,5,6), and 4/12*(4,5,6), which has a similar structure of frequency combination as the amplitude-modulation signal used for the missing fundamental experiment (Schouten et al., 1962). The three notes that make up each chord can match the harmonic template in pitch perception and are integrated again under the SI, defining the pitch of the note two octaves below the respective root note as shown in Fig. 4 (Terhardt, 1978; Takahashi, 2023). However, this missing fundamental has the same chroma as the root note and will not be able to be perceived as an independent tone. We name it the hidden fundamental. The frequency structure of the minor chord cannot be described by three consecutive integers but is represented by 5*(1/6,1/5,1/4) for ACE(Am) and 3*5/2*(1/6,1/5,1/4) for EGB(Em) (Riemann, 1877; Tymoczko, 2010). The three notes in each chord are integrated by accumulation in the subharmonic template matching with missing fundamental into 5 and 15/2, two octaves below 5/4=E and above 15/8=B respectively, are perceived as the hidden fundamental. DFA(Dm) is 1/(8*3)*(27,32,40), which cannot be expressed as a simple integer combination, but is written by 20/3*(1/5.926,1/5,1/4), so 5/3=A will be perceived as the hidden fundamental. BDF(Bdim) does not have a simple integer ratio, but can be written by both consecutive near-integers and reciprocals 3/8*(5,6,7.111) and 45/8*(1/6,1/5,1/4.2188), so the matched fundamental is separated from the three tones, and 3/2=G and/or 45/32=F# will be perceived as the missing fundamentals. The reciprocals of integers appearing in the notes of minor chords are called undertones. They have been known since Rameau's time and are one of the key concepts in Neo-Riemannian theory. Today they are considered a formal structure in music theory and rarely treated as an issue in hearing. From the standpoint of nonlinear dynamics, the generation of subharmonics requires strong nonlinearity and easily transitions into chaos, making it difficult to appear as a perceptible physical phenomenon. However, in the template model of pitch perception, undertones are possible as a perceptual phenomenon in the brain because the subharmonic template matching essentially accompanies the harmonic template matching. Undertones were sounding surely in Riemann's brain as a psycho-phenomenon, even though they could not be heard as a physico-phenomena (Riemann, 1877; Takahashi, 2023).

Next, we focus on the tetrads. As is known, the tetrad called Tristan chord triggered a rethinking of the meaning of tonality in the 20th-century music. It is a variation of the half-diminished chords. The frequency structure of the half-diminished chords, e.g. CEbGbBb, is (1/1,6/5,7/5,9/5)=7*(1/7,1/5.833,1/5,1/3.889). It is approximated by an undertone sequence (1/7,1/6,1/5,1/4). On the other hand, tetrads having the frequency structure of (1/1,5/4,3/2,9/5), e.g. CEGBb, are known as dominant chords. It is rewritten as 1/4*(4,5,6,7.2) and approximated by an overtone sequence (4,5,6,7). Each can be integrated into the missing fundamentals by the harmonic template matching. The integrated tones are the same pitch class as the seventh note with two octaves higher for the half-diminished chord (Bb^{++} for Cϕ7), and the root note with two octaves lower for the dominant chord (C^{--} for Cdom7). The half-diminished chords and the dominant chords are tetrad counterparts to minor and major triads respectively.

In Western music theory, chords have been discussed often as the production of beautiful sounds through the superposition of notes. However, biologically speaking, the meaning of chords should be regarded as a template for Sound Integration in complex tones consisting of overtone and undertone sequences lacking lower harmonics. In other words, chords are merely the reconstruction of the spectral structure by the superposition of musical tones, controlling the timbre (Anderson, 2000; Fineberg, 2000), and the pleasantness of chords is the reward for success in Sound Integration rather than aesthetic beauty.

Chord Progression

Next, template matching is extended in the time domain. Major chords C, G, and F are classified as Tonic, Dominant, and Subdominant, while minor chords Am, Em, and Dm are classified as Tonic, Tonic, and Subdominant respectively. It is noted that Tonic, Dominant, and Subdominant are vectors linking the starting point and the point of arrival rather than a scalar property belonging to each chord in our model. Moving from one key to another and changing the starting note changes the assignment of Tonic, Dominant, and Subdominant. For example, C and G are Tonic and Dominant in C major, while G is Tonic and C is Subdominant in A major. Dominant is formally defined as the frequency multiplication of 1/3 (mod 2) to Tonic and Subdominant is defined as 3 multiplication (mod 2) (Table 4). According to the hierarchy principle, the 3-folding has a greater stability reduction than the 3-accumulation and the 2-power has a smaller SI than the 3-power, so the Dominant transition has a greater stability reduction than the Subdominant transition. Under the hierarchy principle, transitions between two tones are always destabilizing because, with the exception of unison, tone transitions always involve prime-accumulation and prime-folding. However, transitions over three or more tones can cancel out the instability by returning to the first defined reference tone, even if it is once destabilized. The above transitions give a physiological explanation for the cadences in classical chord theory if we redefine destabilization as tension and stabilization as resolution. The fact that Tonic is the "gravitational center" indicates that the first sound given defines the reference point afterward and is the absolute stable point in the perceptual potential.

Our view of chord progressions as mathematical operations is similar to the transformation theory of Lewin et al.(Lewin, 2010; Tymoczko, 2010). Our transformations are represented by a commutative factor group with modulo 2 consisting of T (multiplication by $\underline{\text{T}}$hree) the product of 3/2, F (multiplication by $\underline{\text{F}}$ive) the product of 5/4, and *I* the inversion (reciprocal of each note against the root note). They are interrelated by $P = TI, R = FI,$ and $L = FTI$ for major triads and $P = T^{-1}I, R = F^{-1}I,$ and $L = F^{-1}T^{-1}I$ for minor triads, respectively with the *PRL* transformation used in Neo-Riemannian theory. In the *FTI* transformation, the tonality exchange *I* between major and minor triads is separated, and the translations *F* and *T* on the Tonnetz conserve the tonality of the triads, whereas each *PRL* transformation accompanies the tonality exchange. Dominant is written as T^{-1} and Subdominant as *T*. The *FTI* representation that is derived from the harmonic template gives a physiological meaning to the *PRL* transformations appearing in Neo-Riemannian theory. In the case of minor triads, on the other hand, as Am is assigned to T, there is a correspondence between Em for 3-folding and Dm for 3-accumulation. However, Em is usually identified with T, whereas Dm is identified with S. We now pay attention to Bdim, which has a frequency structure of (15/16,9/8,4/3). It is not able to be written by a combination of simple integers, but is written by 9/16*(1/6,1/5,1/4.219), which is approximated by 9/16(1/6,1/5,1/4). In this case, Bdim is interrelated to Em=3/4*(1/6,1/5,1/4) with a frequency ratio of 3/4. Bdim works as a dominant chord to Em and, therefore, Em can work as a quasi-stable chord that resolves the instability of Bdim. As a result, not only Cm but also Em can have the T function in the minor mode of the diatonic triads. Remember that Bdim can not have any functions in the major mode, as its approximated form is 9/10*(4,5,6), and necessary *F* transformation to interrelate to other major triads.

In summary, consonance is the success of Sound Integration by subharmonic template matching of sounds from multiple sound sources, the degree of consonance corresponds to the combination and the order of the prime-power chain, and tonality can be attributed to whether the notes are made by the prime-accumulation or prime-folding to the root note, chords represent the Sound Integration of stacked sounds through template matching, and chord progressions are the process of destabilization by prime-accumulation and recovery of the stability by returning to the reference point in the perceptual potential.

Neuronal Origin of Hierarchy

We showed that frequency-related perceptual structures, such as pitch and scale, could be organized into a universal hierarchical structure across cultural boundaries within a framework that included not only Western music but also Indian music and jazz. The sensory impact describing the hierarchical nature would be a good explanation of the emotional tendencies produced by music and reproduces the contexts of Western music theory well. The reasons for the preference for consonance and chords have often been reduced to the esthetics that Simple is Beautiful, based on the number-theoretic simplicity that appears in the frequency structure. We emphasize that the emergence of the discrete structure relating prime numbers in the continuous phenomenon of acoustics can be derived from a purely physiological phenomenon based on the survival strategy of SL in ASA, which exploits the detection of small acoustic signals and their associated nonlinearities.

Then, we would like to discuss the origins of the hierarchical nature of the sensory impact. Harmonic template formation and pitch perception were derived as a consequence of adaptation to ASA by SL. We consider the learning and inference process in Sound Integration to be a neuronal reward system. The animals must immediately decide on a subsequent behavior after Sound Integration. Completing matching for all components is not desirable for survival strategies. Template matching of the acoustic signal is performed in parallel for each harmonic component in the acoustic signals. If it is confirmed that the fundamental frequency predicted in part of the harmonic groups is consistent with the other harmonics, matching can be interrupted, and they can take the next behavior immediately. Matching in individual neurons would be rewarding to the neuronal system if the matching succeeds and the matching would be forced to terminate when the total reward exceeds a certain threshold.

The hierarchy of sensory impact in the prime-limit system was considered to consist of three layers: (1) nonlinearity in cochlear amplification (generation of second- and third-order harmonics), (2) matching chains through reinforcement learning (formation of power templates and harmonic templates by $2f_1 - f_2$ coupling), (3) subharmonic matching by folding harmonic templates. The order corresponds to the number of matching chain iterations. The higher the number of chain repetitions, the more noise is mixed in and the lower the correlation, so the higher the order, the lower the responsiveness to matching, and the larger the sensory impact is required to be. Positive orders are accumulation iterations and are learned directly from the input signals. On the other hand, the negative order must calculate the correlation from the fundamental component to the $\frac{1}{n}f$ component by working backward from the correlation from the $\frac{1}{n}f$ component to the $f = n(\frac{1}{n}f)$ component, as there were not so many sounds consisting of tones with rational frequency ratios, at least, before music and language were born. Therefore, the inference of the subharmonics would become harder than that of the harmonics, resulting requirement of higher sensory impact.

The handling of the response strength in perturbation theory needs some caution. The auditory system selectively exploits only second- and third-order nonlinearities. In perturbation theory, if orders of the second- and third-order nonlinearities are equal, then higher-order nonlinearities must arise. In this case, the system easily transitions to a chaotic state due to coupling among the higher-order terms, and the harmonic template cannot exist stably. In nonlinear optics, the generation of higher-order harmonics is described by a chain of third-order nonlinearities via resonances. Similar signals of $f_1 + n(f_1 - f_2)$ are also observed in otoacoustic emission (Kemp, 1979; Bian & Chen, 2008), which should be considered a chain of third-order rather than higher-order nonlinearities (Imasaka, 1989; Takahashi, Mano, & Yagi, 2006). However, since hearing has no particular preference for particular frequencies, the resonant mechanism would not work well. We consider hierarchy in prime numbers to be a phenomenon characteristic of learning in the neuronal system. When using second- and third-order nonlinearities for SI, the third-order nonlinear signal is smaller than the second-order nonlinear signal, so learning both equally would require higher sensitivity of the third-order nonlinear signal relative to the second-order nonlinear signal. On the other hand, in the inference process after learning is complete, the reference fundamental signal and the second- and third-order signals in the input signal have similar signal strength, so a sensitivity correction of the third-order harmonic signal to the second-order harmonic signal is required at the receiving side. A suppressive sensitivity correction would occur in the perception of third-order harmonics through the efferent and afferent interaction between the periphery and the center in learning and inference. As a result, the sensory impact of prime 3 requires a larger value than that of prime 2. Template learning in the fifth harmonic requires signals from both second- and third-order harmonics, and the sensory impact is necessarily greater than matching in the second and third harmonics due to lower learning efficiency. Note that under the $2f_1 - f_2$ matching scheme, the cost of learning the seventh harmonic template is not much different from the fifth. It is probably no coincidence that the Blue Note scale emerged as an extension from minimal 5-accumulation to minimal 7-accumulation.

It has long been understood that music produces emotions. However, it is only relatively recently that the relationship between music and emotion has been treated as a scientific topic (Juslin & Sloboda, 2001). We hypothesized that rewards based on the success of the Sound Integration and gradations of sensory impact generated in the periphery by suppressive sensory impact propagate to emotional tendencies in the perceptual center, giving rise to categorical, dimensional axes and the circumplex structure. We consider that the circumplex structure produced by acoustic signals becomes the basis of the emotions produced by music (Russel, 1980). Needless to say, emotions do not arise from scale, chords, and chord progression alone. Emotions produced by music are expressed as a function of the perceptual potential produced by them. Points themselves in the perceptual potential are not the emotions. Emotions induced by music should be regarded as the product of categorization and integration of lower-order perceptual information input about sound, combined with experiential memories and input stimuli in other modalities at the level of higher-order perceptual information processing. Emotions and music would be considered to be a crossover of affordances provided by various modalities.

Here we point out that rhythm does not appear in our model, even though rhythm is an essential element of music. Pitch, scales, and chords are coded by spatially arranged resonators, the auditory cells. Rhythm, on the other hand, is coded via chaotic dynamics in auditory neural networks. The frequency domains covered by both are almost completely separate. Is it appropriate to expect correspondence between the two? Frequency information is used to integrate sound clusters and map sound sources in the brain under ASA. Rhythmic information, on the other hand, represents the motion of the sound source. Some of the world's music does not have a clear rhythm or melody. Rather, it should be considered that pitch and rhythm are different modalities and that the commonality of the processing object, the acoustic signal, affords the integration of different modalities to become music. The neuroanatomical studies also suggest a model of functional architecture whereby distinct neural circuits for music grouped into pitch organization and temporal organization represent an interactive system of music processing (Peretz & Coltheart, 2003). At the same time when the West was creating theories of harmony, India and the Arabs were enjoying super-rich microtonal music and the Africans were enjoying advanced rhythmic music. Everywhere in the world, different cultures have developed their own rich musical traditions. For example, they sought richness in chords and chord progressions, in rhythm, in modality overlaps with dance and language. The direction of enrichment pursued by each culture may differ, but at least today we can enjoy each other's music. It is important to note that the richness encompassed is not limited to the musical acoustic signals, which Western music has dealt with. Originally, music would have been only part of cross-modal cognitive behaviors. Culture has given each ethno music (including Western music) its own individual taste, but on the other hand, culture might have been masking our perceptual capacities. The contemporary attempt by art to cross modalities such as visual and tactile, as well as auditory, might not be an attempt to create new perceptions by subverting existing perceptions, but rather an act of ancestral reverie, revealing the many innate perceptual abilities that we originally possessed.

Within the extremely vast perceptual network, hearing is particularly well suited for investigating perceptual cross-modalities, as its physics is simple, and it is clearly linked to emotion. Among auditory phenomena, I would like to emphasize that music is a well-designed prototype for examining how the various physiological responses of the body are centrally integrated under survival strategies.

Conclusion

We previously modeled pitch perception by the processes (1) active SA of cochlear, (2) the second- and third-order nonlinearities in hearing, (3) generation of harmonic templates, and (4) Sound Integration by harmonic template matching and attributed the pitch to the success signal of SI for SL in ASA. We have discussed in this paper the processes that produce scales, tonalities, chords, and harmonic progressions by subharmonic template matching. The subharmonic template is generated by folding the harmonic templates back into an octave. Whereas harmonic templates integrate acoustic signals emanating from a single object and perform SI for SL of the object under ASA, subharmonic templates afford the perception of acoustic correlations in multiple acoustic objects. For each of the prime numbers 2, 3, 5, and 7 that make up the harmonic template, the subharmonics hierarchically integrate groups of tones to produce the prime-limit systems of the musical scale. The Pythagorean scale can be represented by a 3-limit system, the just intonation by a 5-limit system, and the 22-note Shruti of Indian music by an extended 5-limit system or 7-limit system. The blue note scale could be regarded as the simplest approximation of the 7-limit system. The sensory impact on prime and order is defined from the hierarchy of the scales. The tonality is defined from the asymmetry in the prime-accumulation and prime-folding with respect to the root note. It is shown that major triads are represented by a group of tones of the Sound Integration at frequency ratios (4, 5, 6), and minor triads are represented under frequency ratios at (1/4, 1/5, 1/6). Furthermore, it is shown that the general tetrads are also an approximation of a group of tones integrated at (4,5,6,7) or (1/4,1/5,1/6,1/7). The hierarchical nature of sensory impact creates a perceptual potential. All of the tone transitions are destabilizing (tense) because they always involve prime-accumulation and prime-folding. Even though tone transition is always destabilizing, it can be stabilized by returning to the tone presented first. Also, returning to the beginning is the only way to stabilize (relaxation). The tone presented first defines the absolute stable point Tonic, as high and low perceptual potentials are relative evaluations. Cadences in chord progression are nothing but dynamics in the perceptual potential. Dominant and Subdominant are defined as transitions with 1/3 and 3 frequency multiplication under modulo 2 respectively. The origin of the hierarchy of the sensory impact is attributed to the difference in sensitivity to second- and third-order harmonics. Hearing needs signal enhancement during learning and signal suppression during inference to compensate for the sensitivity to weak third-order nonlinearity signals. We hypothesized that suppressive control over third-order harmonic signals at the periphery propagated to perceptual information processing in the center, providing suppressive responses to perceptual potentials, producing emotional tendencies in tonality and chords, and giving the basis for the emotions produced by music.

Open practices statement

Neither of the studies reported in this article has been pre-registered. No data other than those presented here are available.

References

Anderson J. (2000). **A provisional history of spectral music.** *Contemporary Music Review* 19(2): 7-22. https://doi.org/10.1080/07494460000640231

Ball P. (2010). **The Music Instinct: How Music Works and Why We Can't Do Without.** Oxford Univ Press on Demand. ISBN-10 : 0199754276

Békésy G. (1949). **The vibration of the cochlear partition in anatomical preparations and in models of the inner ear.** *J. Acoust. Soc. Am.* **21**: 233-245. https://doi.org/10.1121/1.1906502

Bernstein J. G., Oxenham A. J. (2003). **Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number?** *J. Acoust. Soc. Am.* **113** (6):3323-3334. https://doi.org/10.1121/1.1572146

Bian L, & Chen S, (2008). **Comparing the optimal signal conditions for recording cubic and quadratic distortion product otoacoustic emissions**, *J Acoust Soc Am.*, **124**(6):3739–3750. https://doi: 10.1121/1.3001706

Bregman A. S. (1994). **Auditory scene analysis: The perceptual organization of sound.** MIT press. ISBN-10 : 0262521954

Cutting C. B. (2018). **Microtonal Analysis of "Blue Notes" and the Blues Scale.** *Empirical Musicology Review*, **13**(1-2): 84-99.

de Cheveigné A. (2010). **Pitch perception (Chap.4); Oxford Handbook of Auditory Science: Hearing**, Oxford University Press, USA. ISBN-10 : 0199233551

de Cheveigné A. (2004). **Pitch perception models-a historical review.** CNRS-Ircam, Paris, France.

Datta A. K., Sengupta R., Dey N. (2011). **Objective analysis of srutis from the vocal performances of Hindustani music using clustering algorithm.** *EUNOMIOS Open Online J. Theory Anal. Semiot. Music*, 25.

Fineberg J. (2000). **Guide to the basic concepts and techniques of spectral music**, *Contemporary Music Review*, **19**(2): 81-113, https://doi.org/10.1080/07494460000640271

Gill K. Z., & Purves D. (2009). **A biological rationale for musical scales.** *PLoS One* **4**(12):e8144. https://doi.org/10.1371/journal.pone.0008144

Goldstein J. L. (1973). **An optimum processor theory for the central formation of the pitch of complex tones.** *J. Acoust. Soc. Am.* **54**:1496-1516. https://doi.org/10.1121/1.1914448

Harrison P. M. C., & Pearce M. T. (2020). **Simultaneous consonance in music perception and composition.** *Psychological Review*, **127**(2):216–244.

https://doi.org/10.1037/rev0000169

Houtsma A. J. M., & Smurzynski J. (1989). **Pitch of complex tones with many high‐order harmonics.** *J. Acoust. Soc. Am.*, 85: S142. https://doi.org/10.1121/1.2026776

Huron D. (2001). **Is music an evolutionary adaptation?**. *Annals of the New York Academy of sciences*, **930**(1): 43-61. https://doi.org/10.1111/j.1749-6632.2001.tb05724.x

Imasaka T., Kawasaki S., & Ishibashi N. (1989). **Generation of more than 40 laser emission lines from the ultraviolet to the visible regions by two-color stimulated Raman effect.** *Applied Physics B*, **49**(4):389-392. https://doi.org/10.1007/BF00324191

Johnston B. (2010). **"Maximum Clarity" and Other Writings on Music**. University of Illinois Press. ISBN-10 : 0252030982

Juslin P. N., Sloboda J. A. (2001). **Music and Emotion: Theory and Research (Series in Affective Science)**. Oxford University Press. ISBN-10 : 0192631888

Kemp D. T. (1979). **Evidence of mechanical nonlinearity and frequency selective wave amplification in the cochlea.** *Arch. Otol. Rhino. Laryngol.* **224**:37–45. https://doi.org/10.1007/BF00455222

Kubik G. (2008). **Bourdon, blue notes, and pentatonicism in the blues: An Africanist perspective (Chap.1); Ramblin' on my mind: New perspectives on the blues.** Urbana, IL: University of Illinois Press. ISBN-10 : 0252032039

Lewin D. (2010). **Generalized musical intervals and transformations**. Oxford University Press, USA. ISBN-10 : 0199759944

Licklider J. C. R. (1954). **Periodicity pitch and place pitch**. *J. Acoust. Soc. Am.*, **26**:945-945. https://doi.org/10.1121/1.1928005

Maltz R. (1992). **Microtonal techniques in the music of Harry Partch and Ben Johnston**. *Music Research Forum*, **7**:14-37.

McDermott J., Hauser M. (2005). **The origins of music: Innateness, uniqueness, and evolution.** *Music perception*, **23**(1):29-59. https://doi.org/10.1525/mp.2005.23.1.29

Oxenham J. A. (2013). **Revisiting place and temporal theories of pitch**. *Acoust. Sci. & Tech.* **34**(6):388-396. https://doi.org/10.1250/ast.34.388

Peretz, I., & Coltheart, M. (2003). **Modularity of music processing**. *Nature neuroscience*, **6**(7), 688-691. https://doi.org/10.1038/nn1083

Riemann H. (1877). **Musikalische syntaxis: Grundriss einer harmonischen satzbildungslehre**. Breitkopf & Härtel.

Russell J. A. (1980). **A circumplex model of affect**. *Journal of personality and social psychology*, **39**(6): 1161–1178. https://doi.org/10.1037/h0077714

Schouten J. F. (1940). **The Perception of Pitch**. *Phillips Tech. Rev*, **5**:286-294.

Schouten J. F., Ritsma R. J., Cardozo B. L. (1962). **Pitch of the Residue**. *J. Acoust. Soc. Am.*

34:1418-1424. https://doi.org/10.1121/1.1918360

Shamma S., Dutta K. (2019). **Spectro-temporal templates unify the pitch percepts of resolved and unresolved harmonics**. *J. Acoust. Soc. Am.* **145**(2): 615-629. https://doi.org/10.1121/1.5088504

Stephens J. (2017). **The Microtones of Bharata's Natyashastra**. *Analytical Approaches to World Music* **6**(1):1-47. ISSN 2158-5296

Subramanya L., Vasudevan A., Deepak A. S., Ramasangu H. (2022). **Representation Framework for Carnatic Music Melodies Using 22 Shruthis.** *Vidyabharati International Interdisciplinary Research Journal* **14**(1):1-3 ISSN 2319-4979

Takahashi J. (2023). **Power Series Template Matching Model for Pitch Perception.** BIORXIV/2023/525831

Takahashi J., Mano K., Yagi, T. (2006). **Raman lasing and cascaded coherent anti-Stokes Raman scattering of a two-phonon Raman band**. *Optics letters* **31**(10):1501-1503 https://doi.org/10.1364/OL.31.001501

Terhardt E. (1974). **Pitch, consonance, and harmony**. *J, Acoust, Soc. Am.* **55**: 1061-1069. https://doi.org/10.1121/1.1914648

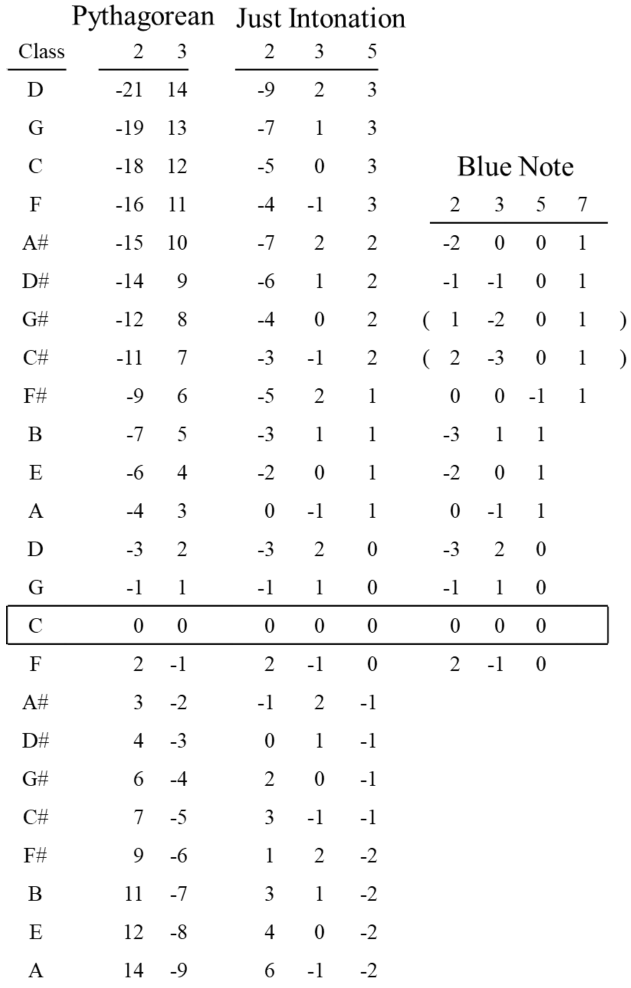
Terhardt, E. (1978). **Psychoacoustic evaluation of musical sounds.** *Perception & Psychophysics*, 23(6), 483-492. https://doi.org/10.3758/BF03199523

Tymoczko D. (2010). **A geometry of music: Harmony and counterpoint in the extended common practice**. Oxford University Press. ISBN-10 : 0195336674

Wightman F. L. (1973). **The pattern-transformation model of pitch**. *J. Acoust. Soc. Am.* **54**: 407-416. https://doi.org/10.1121/1.1913592

Table 1

The combination of prime and order for prime-limit systems and their pitch classes



	Pythagorean		Just Intonation			Blue Note			
Class	2	3	2	3	5	2	3	5	7
D	-21	14	-9	2	3				
G	-19	13	-7	1	3				
C	-18	12	-5	0	3				
F	-16	11	-4	-1	3				
A#	-15	10	-7	2	2	-2	0	0	1
D#	-14	9	-6	1	2	-1	-1	0	1
G#	-12	8	-4	0	2	(1	-2	0	1)
C#	-11	7	-3	-1	2	(2	-3	0	1)
F#	-9	6	-5	2	1	0	0	-1	1
B	-7	5	-3	1	1	-3	1	1	
E	-6	4	-2	0	1	-2	0	1	
A	-4	3	0	-1	1	0	-1	1	
D	-3	2	-3	2	0	-3	2	0	
G	-1	1	-1	1	0	-1	1	0	
C	0	0	0	0	0	0	0	0	
F	2	-1	2	-1	0	2	-1	0	
A#	3	-2	-1	2	-1				
D#	4	-3	0	1	-1				
G#	6	-4	2	0	-1				
C#	7	-5	3	-1	-1				
F#	9	-6	1	2	-2				
B	11	-7	3	1	-2				
E	12	-8	4	0	-2				
A	14	-9	6	-1	-2				

Note. Pythagorean scale is the 3-limit system, Just Intonation is the 5-limit system, and Blue Note scale is the approximation of the simplest 7-limit system.

Table 2

		3 limit			5 limit			7 limit			Datta et.al.		
Note	Calc.	Ratio	Value	Error	Ratio	Value	Error	Ratio	Value	Error	Ratio	Value	Error
Sa	1.0000	1/1	1		1/1			1/1			1/1		
Ri1	1.0467	256/243	1.0535	11.2	256/243	1.0535	11.2	21/20	1.0500	5.4	22/21	1.0476	1.5
Ri2	1.0708	2187/2048	1.0679	-4.7	16/15	1.0667	-6.6	15/14	1.0714	1.1	15/14	1.0714	1.1
Ri3	1.1112	65536/59049	1.1099	-2.1	10/9	1.1111	-0.1	10/9	1.1111	-0.1	10/9	1.1111	-0.1
Ri4	1.1347	9/8	1.1250	-14.9	9/8	1.1250	-14.9	8/7	1.1429	12.4	17/15	1.1333	-2.1
Ga1	1.1731	32/27	1.1852	17.7	32/27	1.1852	17.7	7/6	1.1667	-9.5	27/23	1.1739	1.2
Ga2	1.2056	19683/16384	1.2014	-6.1	6/5	1.2000	-8.0	135/112	1.2054	-0.3	29/24	1.2083	4.0
Ga3	1.2434	8192/6561	1.2486	7.2	5/4	1.2500	9.2	56/45	1.2444	1.5	26/21	1.2381	-7.4
Ga4	1.2711	81/64	1.2656	-7.5	81/64	1.2656	-7.5	80/63	1.2698	-1.8	14/11	1.2727	2.2
Ma1	1.3167	4/3	1.3333	21.8	4/3	1.3333	21.8	21/16	1.3125	-5.5	25/19	1.3158	-1.2
Ma2	1.3486	177147/131072	1.3515	3.8	27/20	1.3500	1.8	27/20	1.3500	1.8	27/20	1.3500	1.8
Ma3	1.3986	1024/729	1.4047	7.4	45/32	1.4063	9.4	7/5	1.4000	1.7	7/5	1.4000	1.7
Ma4	1.4280	729/512	1.4238	-5.1	64/45(729/512)	1.4222	-7.0	10/7	1.4286	0.7	10/7	1.4286	0.7
Pa1	1.4756	262144/177147	1.4798	4.9	40/27	1.4815	6.8	189/128	1.4766	1.1	28/19	1.4737	-2.3
Pa2	1.5100	3/2	1.5000	-11.6	3/2	1.5000	-11.6	3/2	1.5000	-11.6	3/2	1.5000	-11.6
Dha1	1.5730	128/81	1.5802	7.9	128/81	1.5802	7.9	63/40	1.5750	2.1	11/7	1.5714	-1.8
Dha2	1.6126	6561/4096	1.6018	-11.6	8/5	1.6000	-13.6	45/28	1.6071	-5.9	29/18	1.6111	-1.6
Dha3	1.6678	32768/19683	1.6648	-3.1	5/3	1.6667	-1.2	5/3	1.6667	-1.2	5/3	1.6667	-1.2
Dha4	1.7092	27/16	1.6875	-22.2	27/16	1.6875	-22.2	12/7	1.7143	5.1	29/17	1.7059	-3.4
Ni1	1.7586	16/9	1.7778	18.8	16/9	1.7778	18.8	7/4	1.7500	-8.5	7/4	1.7500	-8.5
Ni2	1.8035	59049/32768	1.8020	-1.4	9/5	1.8000	-3.3	9/5	1.8000	-3.3	9/5	1.8000	-3.3
Ni3	1.8690	4096/2187	1.8729	3.6	15/8	1.8750	5.5	28/15	1.8667	-2.2	28/15	1.8667	-2.2
Ni4	1.9139	243/128	1.8984	-14.0	243/128	1.8984	-14.0	40/21	1.9048	-8.2	23/12	1.9167	2.5

Table 2. The frequencies of 23 Shruti clusters calculated by Datta et al. using 150 songs performed by 53 eminent musicians and scholars covering 21 different ragas and approximations by various systems. The assignment of 5-limit system and Datta et al. are from the reference (Datta et al., 2011), 3- and 7-limit systems are ours. The unit of errors is cent.

Table 3.

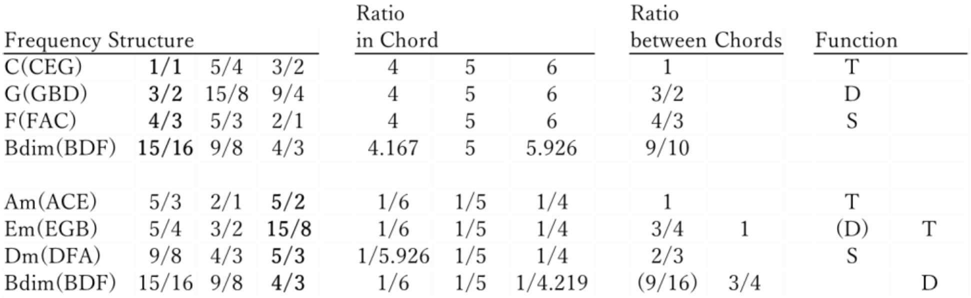
List of combinations of reordered prime and order for 5- and 7-limit systems appearing in Table 2

Class	2	3	5	Class	2	3	5	7
Ma3	-5	2	1	Ma1	-4	1	0	1
Ni3	-3	1	1	Ni1	-2	0	0	1
Ga3	-2	0	1	Ga1	-1	-1	0	1
Dha3	0	-1	1	Dha1	-3	2	-1	1
Ri3	1	-2	1	Ri1	-2	1	-1	1
Pa1	3	-3	1					
Ma4	-9	6	0	Ma3	0	0	-1	1
Ni4	-7	5	0	Ni3	2	-1	-1	1
Ga4	-6	4	0	Ga3	-2	0	1	0
Dha4	-4	3	0	Dha3	0	-1	1	0
Ri4	-3	2	0	Ri3	1	-2	1	0
				Pa1	3	-3	1	0
Pa2	-1	1	0	Pa2	-1	1	0	0
Sa	0	0	0	Sa	0	0	0	0
Ma1	2	-1	0	Ma2	-2	3	-1	0
Ni1	4	-2	0	Ni2	0	2	-1	0
Ga1	5	-3	0	Ga2	-4	3	1	-1
Dha1	7	-4	0	Dha2	-2	2	1	-1
Ri1	8	-5	0	Ri2	-1	1	1	-1
Ma2	-2	3	-1	Ma4	1	0	1	-1
Ni2	0	2	-1	Ni4	3	-1	1	-1
Ga2	1	1	-1	Ga4	4	-2	1	-1
Dha2	3	0	-1	Dha4	2	1	0	-1
Ri2	4	-1	-1	Ri4	3	0	0	-1

Note. The group of MaNiGaDhaRi repeats periodically in both, whereas their indices and position of Pa1 are different from each other.

Table 4

Frequency structure, ratio in and between diatonic triads in the 5-limit system of C major scale.

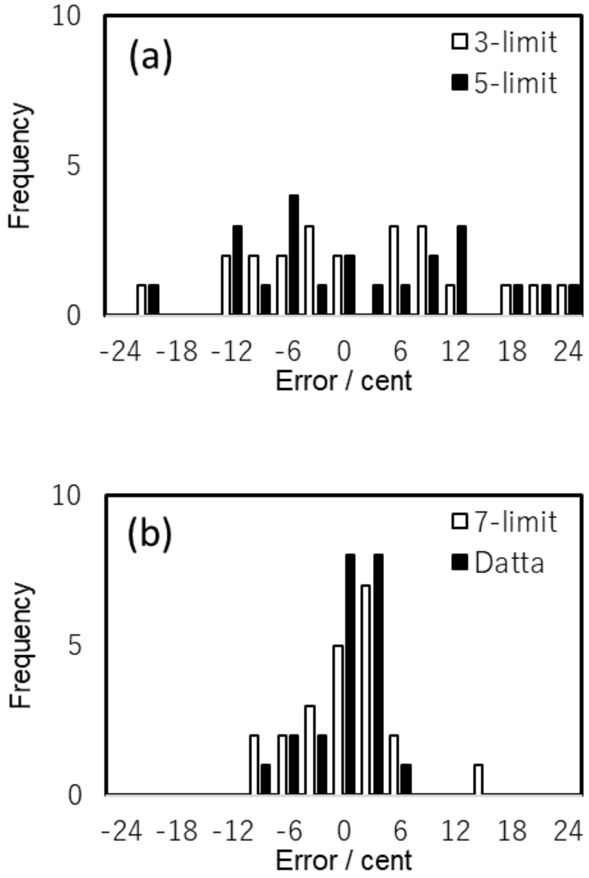


Frequency Structure				Ratio in Chord			Ratio between Chords		Function	
C(CEG)	**1/1**	5/4	3/2	4	5	6	1		T	
G(GBD)	**3/2**	15/8	9/4	4	5	6	3/2		D	
F(FAC)	**4/3**	5/3	2/1	4	5	6	4/3		S	
Bdim(BDF)	**15/16**	9/8	4/3	4.167	5	5.926	9/10			
Am(ACE)	5/3	2/1	**5/2**	1/6	1/5	1/4	1		T	
Em(EGB)	5/4	3/2	**15/8**	1/6	1/5	1/4	3/4	1	(D)	T
Dm(DFA)	9/8	4/3	**5/3**	1/5.926	1/5	1/4	2/3		S	
Bdim(BDF)	15/16	9/8	**4/3**	1/6	1/5	1/4.219	(9/16)	3/4		D

Note. Major and minor triads have the frequency ratio of (4,5,6) and (1/6,1/5,1/4), resulting in integration into the hidden fundamental, the pitch classes of which are shown in boldface. The chord functions are defined by the 3-accumulation or 3-folding between chords. Bdim has an approximation to interrelate to Em by 4/3 multiplication, which enable for Bdim to work as a dominant to Em, resulting Em as a quasi-stable chord resolving the instability of Bdim.

Figure 1

Distribution of error for 22 Shruti (Sa is not included.)

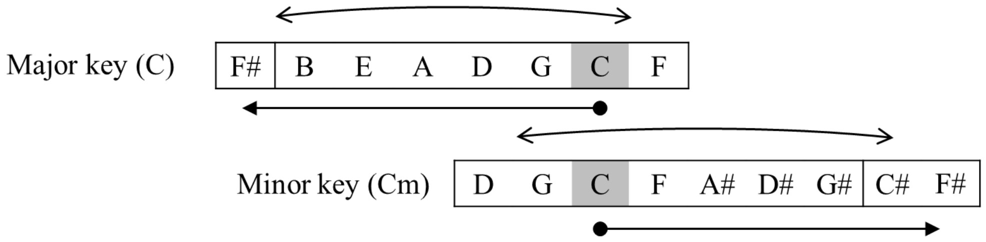


Note. (a) using the assignments by 3- and 5-limit just intonation. (b) using the assignments by Datta et.al. and 7-limit just intonation (ours).
Data are taken from Table 2.

Figure 2

Position of notes in major and minor scales in the 3-limit system (Pythagorean scale)

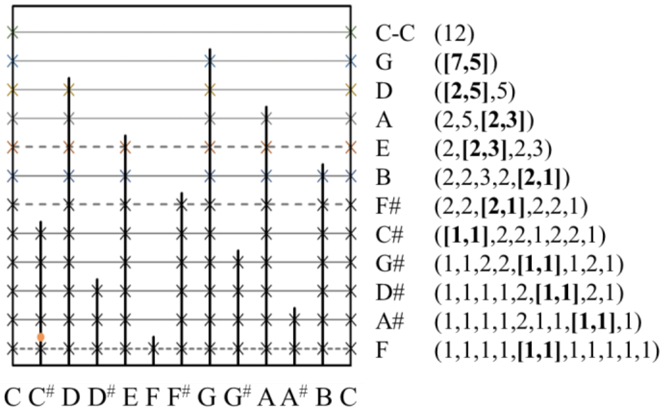
Class	A#	D#	G#	C#	F#	B	E	A	D	G	C	F	A#	D#	G#	C#	F#
3	10	9	8	7	6	5	4	3	2	1	0	-1	-2	-3	-4	-5	-6
2	-15	-14	-12	-11	-9	-7	-6	-4	-3	-1	0	2	4	5	7	8	10



Note. Major scale is defined as the consecutive seven notes of 3-accumulation from C, where the left-most F# is replaced by an adjacent F to the right of C. Minor scale is defined as the consecutive seven notes of 3-folding from C, where the right-most C# and F# are replaced by adjacent D and G to the left of C.

Figure 3

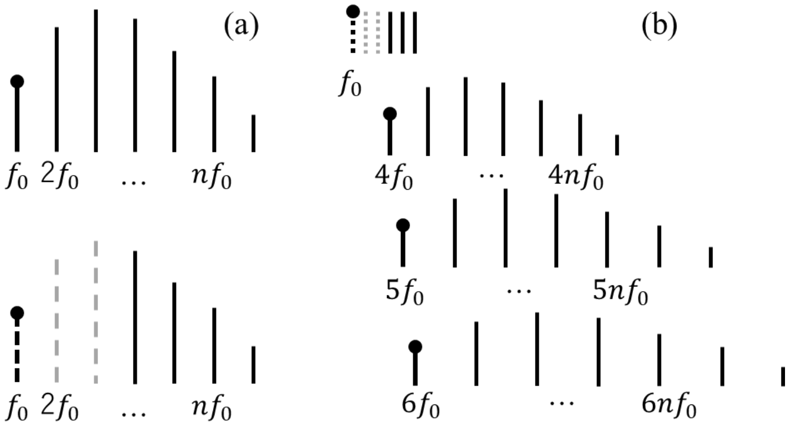
Octave division by addition of note under prime-accumulation



Note. The added note (except D) divides the widest interval and at the same time equalizes the distribution of intervals. The numbers in round brackets indicate the intervals among the notes, and the square bracket indicates the division of the subinterval by the addition of a note. The number distribution is more uniform at 5(E), 7(F#), and 12(F) notes than others.

Figure 4

Spectral structure of pitch and chord perception



Note. (a) pitch perception: natural integer overtones are integrated into the fundamental frequency f_0, the pitch, even if some of the lower harmonics are absent. (b) chord perception: a major chord is composed of three musical tones with the fundamental frequencies of $4f_0$, $5f_0$, and $6f_0$, which are integrated into the missing fundamental f_0.