OPEN ACCESS Freely available online

E. Deruty, P. Arbez-Nicolas, and D. Meredith, "Methods for Pitch Analysis in Contemporary Popular Music: Deviations From 12-Tone Equal Temperament in Vitalic's Work," *J. Audio Eng. Soc.*, vol. 73, no. 5, pp. 256–281 (2025 May).
https://doi.org/10.17743/jaes.2022.0203.

Methods for Pitch Analysis in Contemporary Popular Music: Deviations From 12-Tone Equal Temperament in Vitalic's Work

EMMANUEL DERUTY,[1,3,*] *AES Associate Member*, **PASCAL ARBEZ-NICOLAS,**[2] **AND DAVID MEREDITH**[3]
(emmanuel.deruty@sony.com) (dave@create.aau.dk)

[1]*Sony Computer Science Laboratories, Paris, France*
[2]*Citizen Records, Dijon, France*
[3]*Department of Architecture, Design and Media Technology, Aalborg University, Denmark*

Aims: Twelve-tone equal temperament (12-TET) is recognized as the modern standard tuning for music. However, contemporary popular music may exhibit significant deviations from this framework. This study investigates such deviations in the music of the electronic musician Vitalic and others. **Methods:** The study examines relations between signal features and perceived pitches. The artist's involvement ensures that the analysis focuses on aspects relevant to the music. **Results:** Deviations from 12-TET can be observed as a result of 1) tuning of quasiharmonic tones outside of 12-TET, 2) audible mistuned partials within an otherwise quasiharmonic context, and 3) subsets of partials perceived as distinct pitches outside 12-TET. Examples from other artists suggest that the use of pitches outside 12-TET is not limited to Vitalic's music. **Conclusions:** The deviations from 12-TET are deliberate and pervasive. They are often linked to the acoustic properties of the tones, suggesting 1) a continuum between timbre and pitch and 2) the concept of a "resulting pitch" analogous to the idea of "resulting harmony" in Renaissance polyphonic music. Such conclusions challenge the traditional view of musical pitch.

0 INTRODUCTION

0.1 Musical Pitch in the Literature

The consideration of pitch in a musical context can be broadly categorized into two main areas. The first area involves 1) the identification of pitch in music, through either a) algorithmic analysis or b) the human auditory system; and 2) the organization of pitches within the musical discourse.

With regard to 1a above, algorithmic analysis spans a variety of fields and methods. *Diachronic studies* have used chroma features to extract pitch values [1, 2]. *Automatic music transcription* employs signal processing techniques, such as nonnegative matrix factorization or neural network-based approaches, with tones carrying pitch generally although not necessarily considered to be harmonic [3]. *Monophonic pitch tracking* is framed as an f_0 estimation task, where the tones are always considered harmonic [4, 5]. The assumption that tones generating a sense of pitch are harmonic may stem from acoustic instruments being regarded as linear harmonic systems, even though this claim does not hold under closer examination [6].

With regard to 1b above, the study of how the human auditory system interprets pitch from a signal falls within the domain of psychoacoustics. This field identifies two primary mechanisms for pitch extraction by the auditory system, applicable to both harmonic and inharmonic tones: spectral and temporal modeling [7]. Broadly, spectral modeling relates pitch perception to the frequency spectrum, whereas temporal modeling relates it to autocorrelation.

With regard to 2 above, musicologists define pitch as the quality of a sound that determines its position within a scale [8]. Although Western 12-tone equal temperament (12-TET) is widely recognized as the modern standard [9], other tuning systems have been used. Earlier Western temperaments involved 12 nonequal divisions of the octave [10]. Other systems may use a different number of equal divisions, as in traditional Thai music [11]. Twentieth-century systems, such as extended just intonation, may involve a higher number of degrees [12]. Musical pitch may also exist outside discrete tuning, with composers such as Xenakis incorporating continuous pitch evolutions [13].

*To whom correspondence should be addressed, email: emmanuel.deruty@sony.com

The diachronic studies cited above, along with the automatic music transcription literature, adhere to the 12-TET convention. Automatic music generation may reflect this convention, as seen in chord conditioning approaches [14].

Overall, the literature conceptualizes musical pitch predominantly in relation to the f_0 of harmonic tones, generally organized within 12-TET. Alternative configurations are typically regarded as deviations from the general framework.

0.2 Observations From Popular Music

The motivation for the present study arises from observation. In contemporary popular music as defined by Deruty et al. [15], pitches can be heard to deviate from 12-TET. Among many examples, one can cite Diplo's song "Express Yourself" [16], productions by Hyper Music [17], and several works by Vitalic, whose music forms the primary focus of this paper. These examples are mainstream productions from artists with wide audiences who license their work for use in films, television series, and commercials. As a result, they likely reflect current practices of the music industry.

The original aim of the study was to confirm these observations by linking perceived pitches that deviate from 12-TET to f_0 values. However, signal analyses revealed additional insights: in Vitalic's music and in Hyper Music's works [17], the signal properties associated with perceived pitches are generally not f_0 values. All observed tones exhibit some degree of inharmonicity. Perceived pitches may be associated with 1) the f_0 of the closest harmonic tone in quasiharmonic tones, and 2) mistuned partials in an otherwise quasiharmonic context, and 3) subsets of overtones perceived as individual pitches.

Vitalic's music is not culturally isolated. It is produced using widely available technology and draws its influences from previous artists such as Daft Punk. Although the study should be extended to the production of other artists to gain validity, it suggests that there are at least some music genres in which deviations from 12-TET are the norm. Additionally, the consensus according to which musical pitch can generally be described as close to the f_0 of (quasi)harmonic tones may be questioned.

0.3 Vitalic

Vitalic is the stage name of French electronic musician, Pascal Arbez-Nicolas, of whom Phares provides a comprehensive biography [18]. Since 2001, Vitalic has authored several albums, EPs, and singles. His music has found its way into commercials, video games, movies, and television series. Vitalic's style blends two approaches. On the one hand, he creates "dance floor" tunes—dry, rhythmic pieces with few notes, such as "And it Goes Like" [19]; on the other hand, he also produces more "nostalgic" songs with "more notes" (Vitalic's words), such as "Eternity" [20]. His discography is a mix of these influences. A key element in his music is the deliberate introduction of *uncertainty*. For instance, in "Nozomi"[21],

1. The kick and snare drums keep a steady beat, but the bass rhythm is deliberately unquantizsed, creating rhythmic uncertainty;
2. In the bass track, it is difficult to discern whether one or two pitches are present at any time, and the octave may be uncertain, leading to ambiguity in the number of pitches; and
2. The pitches of a repetitive keyboard part (between 1'10 and 3'10) do not appear to follow a fixed scale; instead, they never seem to be exactly the same.

In this study, the focus is on the third aspect. Vitalic traces his preference for elements that do not strictly conform to a fixed scale back to his experience of listening to Jean-Michel Jarre's "Equinoxe" [22] on a sun-damaged vinyl record during his youth, initially believing it to be the original version. As he grew up, he identified and appreciated similar traits in other music, including the film scores of François de Roubaix, the vocal performances of William Sheller, and Giorgio Moroder's works, such as Sparks' "Tryouts for the Human Race" [23]. In his own music production, Vitalic claims that "all [the] sounds are always out of tune" [24], as part of a process to humanize machine-generated sounds.

0.4 Pitch in Vitalic's music

In this study, the aim is *not* to determine the pitch(es) perceived when listening to sounds, but rather to describe aspects of the signal that may relate to pitch perception. In the examples from Vitalic's music, deviations from 12-TET arise from various factors, including the following:

- *Subsets of partials* being perceived as unitary percepts, each subset with its own pitch, making tuning dependent on the partials' frequencies (see SEC. 2.1);
- Two perceived pitches, each with a different tuning, resulting from a combination of *temporal modeling* and the *fundamental frequencies* (see SECS. 1.3 and 2.2);
- Pitches being mainly influenced by fundamental frequencies fitting a different temperament (see SEC. 2.3);
- Conversely, pitches being mainly related to *frequency differences between partials* (linked to temporal modeling), also aligning with another temperament (see SEC. 2.4);
- Deviation from equal temperament being further accentuated by a *glissando* (see SEC. 2.5);
- Deviation from equal temperament being emphasized by *reverberation modes* (see SEC. 2.6); and
- A minor triad from detuned *individually-heard partials* resulting in a chord that does not align with equal temperament (see SEC. 2.7).

In all examples, pitches are never static, and the partials' frequencies contributing to pitch perception evolve, often independently. The deviations from 12-TET may not form a consistent tuning system because different octaves of the

Table 1. Extracts of Vitalic songs used in SEC. 2.

Album	Song	Reference	Timing	Section
OK Cowboy	"No Fun"	[26]	0'12 to 0'14	2.1
Rave Age	"Next I'm ready"	[27]	0'52 to 0'59	2.2
Dissidænce Episode 1	"Cosmic Renegade"	[28]	2'54 to 3'02	2.3
Voyager	"Levitation"	[29]	2'38 to 2'42	2.4
OK Cowboy	"Polkamatic"	[30]	0'00 to 0'19	2.5
Voyager	"Lightspeed"	[31]	0'55 to 1'10	2.6
Dissidænce Episode 2	"Winter Is coming"	[32]	0'47 to 0'50	2.7

same pitch class may be tuned differently (see SEC. 2.4). Additionally, multiple temperaments may coexist simultaneously (see SEC. 2.2).

0.5 Structure

In SEC. 1, the methodological and theoretical frameworks are introduced. Next, the analyses themselves are presented in SEC. 2, with each analysis being structured as follows: 1) context and motivation, 2) the analysis, and 3) a conclusion. The results of these analyses are summarized in SEC. 2.8. Finally, in SEC. 3, music from other artists is examined to show that deviations from 12-TET are not unique to Vitalic. Supplementary material with audio examples and detailed analyses is available at http://vnet-suppl-mat.s3-website.eu-west-3.amazonaws.com/.

1 ANALYTICAL METHODS

1.1 Analytical Process

Seventy-two songs originating from seven Vitalic records released between 2005 and 2023 were imported into a Pro Tools session and subjected to critical listening. The focus of the listening was *pitch*—not the high-level organization of pitches, such as chord sequence identification, but lower-level criteria, such as the complexity and nature of the tones carrying pitch or the compliance of pitches to 12-TET. Observations were sorted into the following six nonexclusive categories:

1. Presence of high-complexity instrumental tones;
2. Presence of complex vocal parts (e.g., use of vocoder);
3. Pitches not consistent with 12-TET;
4. Presence of continuous pitch trajectories;
5. Presence of low pitch-strength instruments [25] (e.g., drums); and
6. Presence of flanger/chorus resulting in additional pitch.

In this study, the focus is on category 3 while incorporating aspects of category 4. Fifty-three of the seventy-two songs were found not to comply with 12-TET. As a result, although the study's conclusions do not apply to all of Vitalic's productions, they reflect most of his discography. The primary criterion for selecting the seven songs analyzed in SEC. 2 (listed in Table 1) was the diversity of features contributing to the perception of pitch. Specifically,

1. In SEC. 2.1, a single tone in the bass part is considered that evokes two simultaneous pitches associated with different subsets of its partials;
2. In SEC. 2.2, a single bass-part tone is considered that evokes two simultaneous pitches, one arising from temporal modeling and one associated with the f_0;
3. In SEC. 2.3, the authors consider a case where a single bass-part tone evokes multiple pitches associated with different subsets of the partials, while simultaneously a tone in the lead-synth part evokes two pitches resulting from temporal modelling combined with the f_0;
4. In SEC. 2.4, lead-synth part that features wide, continuous, pitch trajectories and evokes multiple pitches resulting from temporal modeling and f_0 is considered;
5. In SEC. 2.5, a lead-synth part that evokes pitches resulting from both temporal modeling and f_0 simultaneously with a glissando is considered;
6. In SEC. 2.6, another lead-synth part in which pitches result from variable contributions of temporal modeling and f_0 is considered; and
7. In SEC. 2.7, a case where a "chord" is generated from partials whose tuning changes over time, while a lead-synth produces tones that evoke multiple pitches through a combination of temporal modeling and f_0 is considered.

SEC. 1.2 provides the definitions of the terminology (e.g., tone, partial, f_0), and SEC. 1.3 clarifies the notion of temporal modeling. The analyses in SEC. 2 include transcriptions and signal analyses. The transcriptions into Western classical notation are approximate and do not account for deviations from 12-TET.

Signal analyses were primarily performed on weighted audio (see SEC. 1.4), and derived from source separation when applicable (see SEC. 1.5). Each tone that formed the specific focus of an analysis was found to be inharmonic. The analyses are based on the short-time Fourier transform (STFT) and short-time autocorrelation (STAC). Because of the complexity of the content, tracking of partials can be a challenging and time-consuming *annotation* process, where algorithmic suggestions are manually corrected. Only the lower partials were tracked because higher

Table 2. Extracts of songs from other artists used in SEC. 3.

Artist	Song	Reference	Timing	Section
Pink Floyd	"Shine On You Crazy Diamond," part 1	[37]	2'10 to 2'34	3.1
Kanye West	"On Sight"	[38]	2'01 to 2'03	3.2
Primaal	"Whomp"	[39]	0'48 to 1'05	3.3

partials are harder to track, and previous studies have shown that the second to fifth partials are dominant in determining the pitch of complex tones [33–36].

To demonstrate that features leading to deviations from 12-TET in Vitalic's music can also be found in the work of other artists, we include analyses of extracts from three songs by other musicians. These are listed in Table 2.

1.2 Terminology: Tones, f_0, Partials, and Harmonics

Regarding tones, the following terminology is used. One key motivation for this terminology is its applicability to the audio content analyzed in this paper.

1. *Partial* refers to the spectral representation of a single periodic sine wave, corresponding to what Helmholtz describes as a *constituent* or *partial tone* [40].
2. A *harmonic complex tone* is an ensemble of partials that are integer multiples of a fundamental frequency as defined by Helmholtz as a *compound* [40]. The fundamental frequency, denoted as f_0, can still be determined as the greatest common divisor of the partials' frequencies, even if the partial corresponding to f_0 is absent.
3. A *complex tone* (or *tone*), is defined to be an ensemble of discrete partials that share a "common fate" [41]. Typically, the frequencies and/or amplitudes of these partials change in parallel, for instance, when the partials start and stop simultaneously. Common fate as a characteristic of tones was noted by Rasch and Plomp [42]. This definition permits some ambiguity, although no examples in this paper raise questions about tone identification. Such *tones* may evoke more than one pitch, making the convention incompatible with Moore's [43], for which a tone corresponds to one single pitch. Additionally, whereas the definition may encompass tones that do not evoke pitch, all examples in this paper are associated with pitch perception.
4. An *inharmonic tone* is a tone that is not harmonic. An inharmonic tone can still display regularity in terms of where its partials are positioned. In this paper, all tones are inharmonic but exhibit such regularity. Following Rasch and Plomp [42], the notation "f_0" is used in the case of inharmonic tones evaluated as the fundamental frequency of the least deviating harmonic series.
5. An *overtone* or *upper partial* refers to any partial in a tone other than f_0. This aligns with Helmholtz's definition of upper partial tones [40].
6. Helmholtz defines *harmonics* as the sinusoidal components of a harmonic complex tone, where the n^{th} harmonic has a frequency n times the fundamental [40]. We extend the concept to inharmonic tones with regular partial positioning, where the n^{th} harmonic's frequency is near the n^{th} multiple of the f_0 of the least deviating harmonic series.

Fig. 1 illustrates some of the terminology with a bass tone from "Poney Part 1" [44].

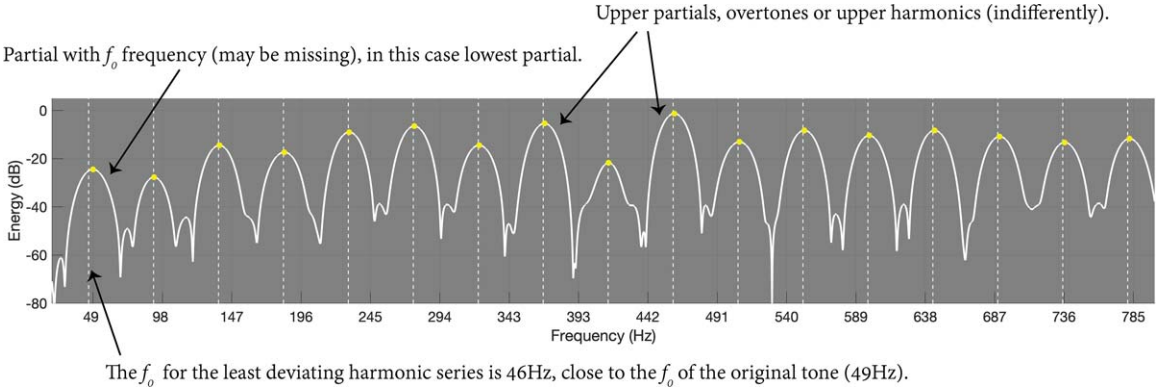


Fig. 1. "Poney Part 1," 0'53 to 0'55, Fourier transform for one frame. Solid white curve shows power spectrum for the tone. Yellow dots indicate peak positions. The x-axis grid is set on the multiples of the first peak. Dashed white lines indicate frequencies of the partials corresponding to the least-deviating harmonic series.

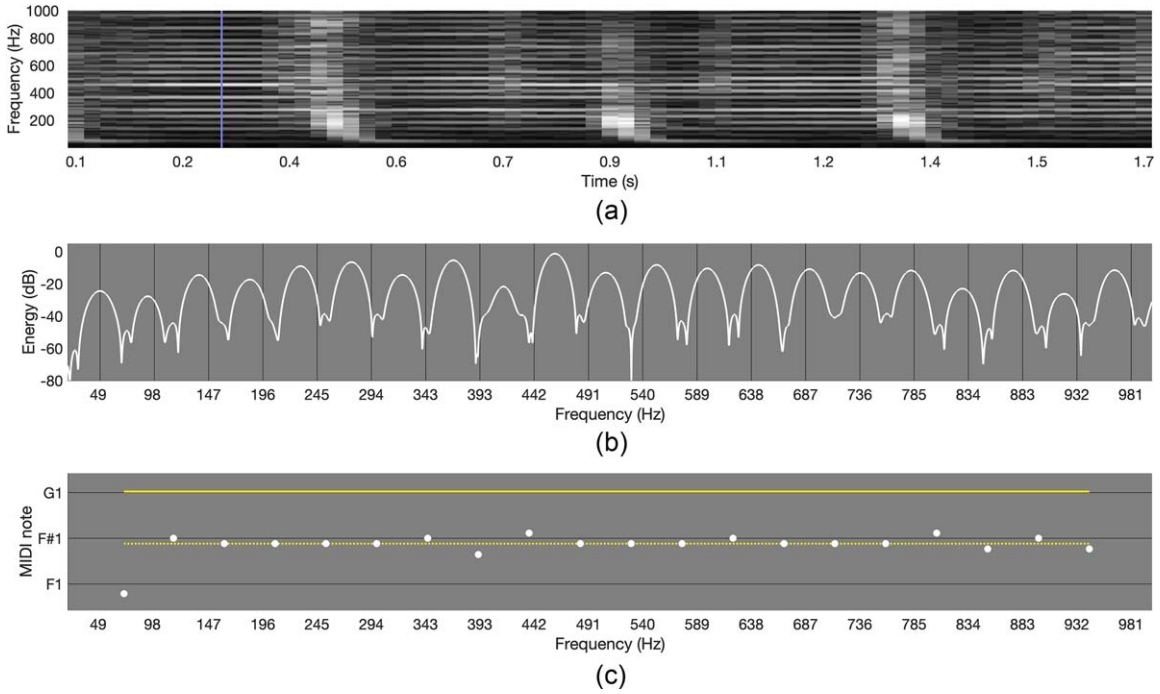


Fig. 2. "Poney Part 1," 0'53 to 0'55. See supplementary material SEC. A for the same figure with synchronized audio. (a) STFT with weighted audio. (b) Fourier transform of frame shown by the vertical blue line in (a). The x-axis grid is aligned with multiples of the lowest partial. (c) Scatter plot showing frequency differences between the peaks in (b). Solid yellow line indicates first partial's frequency. Dotted yellow line indicates frequency value derived from autocorrelation.

1.3 Spectral and Temporal Modeling of Pitch Perception

Yost [7] distinguishes between *spectral* and *temporal* modeling of pitch perception, whose main proponents are Goldstein [45] and Licklider [46], respectively. In spectral modeling, pitch is extracted from the spectral components *resolved* by the auditory periphery (i.e., whose frequency difference is above the critical bands' width). In temporal modeling, pitch is derived from autocorrelation. The need for the simultaneous use of both approaches in analyzing Vitalic's music is illustrated in Figs. 1 and 2. The tone is not harmonic, and the frequencies derived from temporal modeling differ from the f_0. In this case, the perceived pitch is close to F♯1, corresponding to temporal modeling.

In this study, spectral modeling is used in a simplified form to link perceived pitches with the f_0 or an individually audible upper partial. Temporal modeling is applied when 1) connecting perceived pitches to the highest autocorrelation peak excluding zero lag [47] or 2) analyzing frequency differences between consecutive partials, thus referring to the *difference tone* or *residue* according to Schouten [48] and subsequent publications [49].

1.3.1 Inharmonic Tones

When tones are inharmonic, the highest autocorrelation peak (excluding zero lag) may not correspond to any perceived value, making it necessary to analyze the frequency differences between consecutive partials. For example, Fig. 3 compares the highest autocorrelation peak (excluding zero lag) with the frequency difference between partials for a series of increasingly inharmonic tones. In these tones, the f_0 remains at 220 Hz (A3), whereas the overtones' frequencies are multiples of increasing frequencies from 220 (A3) to 246.94 Hz (B3). A listening test indicated that when the overtones are multiples of 246.94 Hz, 19 out of 20 participants perceive A3 and B3 simultaneously—the f_0 and the difference tone. The highest autocorrelation peak (excluding zero lag) does not correspond to any perceived pitch.

When inharmonic tones evoke a single pitch, Rasch and Plomp suggest that this pitch may be approximated by the fundamental of the closest harmonic series [42]. Here, this is approximated by finding the f_0 that minimizes the mean distance to the inharmonic tone's partials. In such cases, pitch analysis involves all partials, not just the lowest. Examples from all analyses in SEC. 2 except SEC. 2.2 fall into this category.

1.3.2 Harmonic Tones

For harmonic tones, the highest autocorrelation peak (excluding zero lag) generally matches the frequency difference between partials. For harmonic tones with odd harmonics only, the first autocorrelation peak appears an octave below the perceived pitch, aligning with the frequency difference between partials. The observation is further detailed in SEC. 2.2, in supplementary material SEC. C, and by De Boer [49].

1.4 Psychoacoustic Weighting

Signal analyses in this study incorporate equal-loudness contour weighting, recognizing that humans are not equally sensitive to all frequencies [50]. Various models exist that

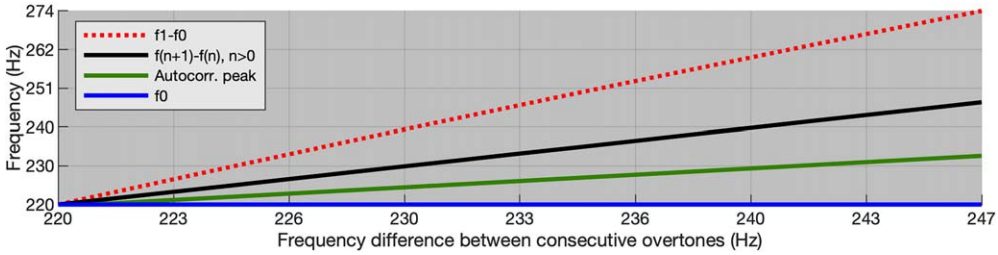


Fig. 3. f_0, highest autocorrelation peak (excluding zero lag), frequency difference between overtones, and frequency difference between the first two partials. The x-scale shows the frequency of which overtones are integer multiples. The leftmost tone is harmonic, $f_0 = 220$ Hz.

weight the power spectrum to reflect perception [50–52]. One such model is International Organization for Standardization Standard 226:2023 [53]. We apply the Standard 226:2023 50-phon equal-loudness contour before spectral analysis to produce results that better reflect what listeners actually hear.

1.5 Entire Mix or Source Separation

The analyses are based either on the original song version or on audio files obtained via source separation using the X-UMX algorithm [54]. Source separation aids in focusing on specific tracks, such as in "Next I'm Ready" (SEC. 2.2) and "Levitation" (SEC. 2.4). However, its use is limited when simultaneous tones are grouped together, such as in "Cosmic Renegade" (SEC. 2.3), "Polkamatic" (SEC. 2.5), and "'Winter Is Coming" (SEC. 2.7).

Note that Vitalic's bass tracks often differ from a traditional electric bass, leading to it being categorized differently by the algorithm. In "Lightspeed" (SEC. 2.6), for example, the lower bass spectrum is classified under the "bass" category, whereas the upper bass is categorized as "other." Similarly, the lower part of the keyboard is assigned to the "bass" category, whereas the upper part appears in the "other" track. The source separation output was carefully examined to ensure that all partials corresponding to the original part were present in the selected output tracks.

1.6 Analyses Were Reviewed by Vitalic

Vitalic reviewed and approved all the results in this paper before submission. The analysis results were regularly discussed with Vitalic throughout the research and writing process. These discussions focused on three main points:

1. *Validation*. Were the analysis results relevant? Did they highlight important aspects of the musical discourse?
2. *Production methods*. What instrumental processes led to these results? Often, the process was too complex to trace entirely because the pitches may result from a combination of oscillator tuning, distortions, and reverberations.
3. *Artistic intent*. Beyond technical aspects, what was the purpose of diverging from 12-TET?

2 ANALYSES

This section presents analyses of extracts from Vitalic's "No Fun," "Next I'm Ready," "Cosmic Renegade," "Levitation," "Polkamatic," "Lightspeed," and "Winter Is Coming." All analyses follow a consistent structure. First, the motivations for selecting each extract are explained. Next, the analysis is conducted, aiming to link the perceived pitches to specific signal properties. Each analysis is supplemented by additional content available in the supplementary material. Finally, we summarize the signal characteristics that result in deviations from 12-TET.

2.1 "No Fun," Bass

2.1.1 Context and Motivation

"No Fun" [26] is an upbeat, mostly instrumental, electronic piece. The main pitched element is a synth-based track running throughout, following a consistent rhythmic pattern. This track is called the "bass" because it functions like an electric bass in rock or pop. It involves the perception of two homorhythmic melodic lines. Pitches in the top line sound "out of tune," and the two lines do not seem in tune with each other. For instance, the A4 in the transcription shown in Fig. 4(a) sounds lower than the simultaneous A3. The analysis links these perceived pitches to characteristics of the signal and offers explanations for why they may sound out of tune.

2.1.2 Analysis

Fig. 4(b) shows the STFT for the corresponding audio extract along with the f_0 values of the transcribed notes, quantized to the nearest partial. The STFT contains only regularly spaced partials, suggesting a sequence of unique tones ("carrier" tone). The part was made with inharmonic tones sampled from the Roland VariOS. Fig. 4(c) displays the frequency distribution of partials closest to the transcribed notes' f_0s. The frequencies do not comply with equal temperament.

Fig. 4(d) focuses on beat 2, second eighth-note (A3 and A4). The tone's f_0, derived from the frequency difference between the first nine partials, is A1 − 29 cent (ct). The audible pitches are carried by subsets of partials. The f_0 of the closest harmonic tones (see SEC. 1.3) neighboring the transcribed pitches are A3 + 29 ct and A4 − 27 ct. The octave is narrower than equal or just temperament.

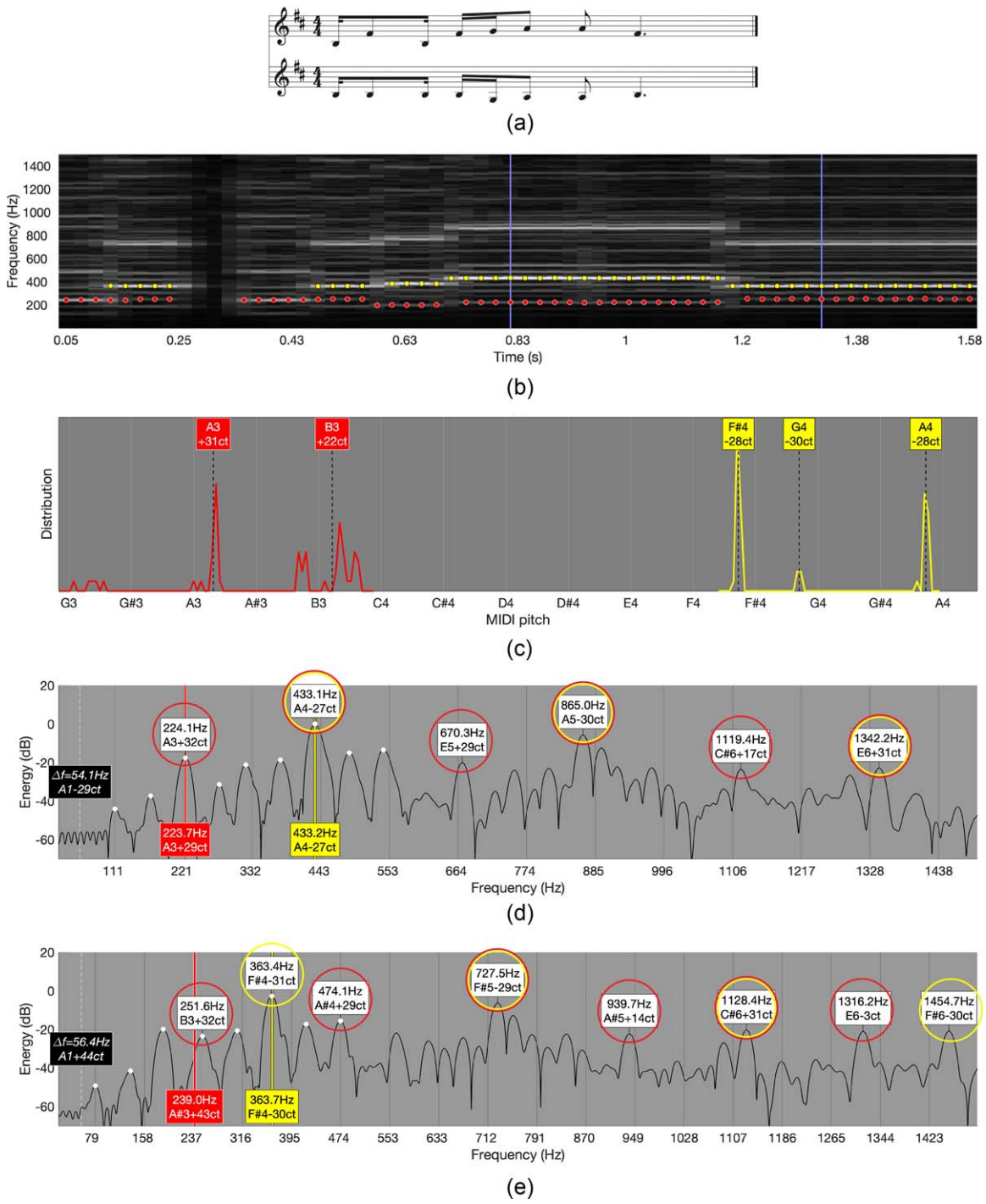


Fig. 4. "No Fun," 0'12 to 0'14. See supplementary material SEC. B for the same figure with synchronized audio. (a) Approximate transcription. (b) STFT with weighted audio. Red and yellow scatter plots indicate frequencies corresponding to the bottom and top staves, respectively, quantized to the closest partial. Blue lines indicate frames featured in (d) and (e). (c) Distribution for the frequencies in (b). (d) Fourier transform at position indicated by left-hand vertical blue line in (b). Red and yellow circles indicate the partial subsets corresponding to bottom and top staves, respectively. Red and yellow text boxes give f_0 of closest harmonic tone. Black text box gives frequency difference between partials indicated by white dots. (e) Same information as (d), but for right-hand vertical blue line in (b).

Fig. 4(e) focuses on beat 4 (B3 and F♯4). The tone's f_0, derived from the frequency difference between the first eight partials, is A1 + 44 ct. The audible pitches are carried by subsets of partials. The f_0 of the closest harmonic tones neighboring the transcribed pitches are A♯3 + 43 ct (or B3 − 57 ct) and F♯4 − 30 ct. The fifth is narrower than equal or just temperament. A comparison with A3 and A4 reveals that the melodic lines, individually, do not comply with equal temperament. In this example, as in all subsequent examples in this study, the deviation from equal temperament is implemented manually, without using temperament settings available in various audio products. See

Table 3. "No Fun," 0'12 to 0'14. 'F_0', f_0 deviations from A♯1, 440-Hz tuning.*

Group	B3	B3 + F♯4	B3	B3 + F♯4	G3 + G4	A3 + A4	B3 + F♯4
F_0	+14 ct	−67 ct	+18 ct	+43 ct	−2 ct	−129 ct	−56 ct
Bottom	−34 ct	−57 ct	−35 ct	−55 ct	+22 ct	+29 ct	−57 ct
Top	/	−34 ct	/	−34 ct	−30 ct	−27 ct	−30 ct

* "Bottom" and "Top" are deviations from the notes transcribed in Fig. 4(a) bottom and top staves.

supplemental material SEC. B, for the analysis of the entire extract leading to the results in Table 3.

2.1.3 Conclusion

The perceived pitches are linked to subsets of partials from a single tone with a much lower fundamental frequency. The deviation of these pitches from 12-TET arises from the positions of the partials within the carrier tone. These observations suggest a continuum between pitch and timbre, where the perceived pitches and their tuning are determined by the positions and amplitudes of the partials within a single tone that is not directly perceived as such.

2.2 "Next I'm Ready," Bass

2.2.1 Context and Motivation

"Next I'm Ready" [27] is a dance piece with a four-on-the-floor bass drum track and a prominent bassline. The perceived timbre of the bass changes throughout the piece, with one notable shift occurring at 0'52. The section between 0'52 and 1'20 is sparse, featuring only bass, drums, and vocal patterns. The analysis is motivated by the perception that the intervals between the notes in the bass part sound out of tune, and by the perception of each note as dissonant. In this analysis, the perceived pitches in the bass track are linked to specific signal characteristics and offer explanations as to why the intervals sound out of tune and why the notes themselves may be perceived as dissonant.

2.2.2 Analysis

Fig. 5(a) illustrates how a detailed listening of the bass track from 0'52 to 0'59 results in a transcription as two parallel melodic lines separated by an octave. The pitches in the lower staff may not be audible when using speakers with limited bass response. The pitches in the higher staff exemplify Vitalic's claim that he likes to create melodies within the bass [24; 3'00]. Fig. 5(b) shows the STFT of the weighted audio for the bass track obtained through source separation. The tone primarily features odd-numbered harmonics.

The identification of two pitches from one tone may be related to the case described by Yost, where two pitches can be linked to a tone containing only odd-numbered harmonics and a missing fundamental [7]. A difference between the bass track in "Next I'm Ready" and Yost's example is that the fundamental in the bass track's tones is relatively weak rather than entirely absent.

Fig. 6(a) shows the maximum autocorrelation value for each frame. Fig. 6(b) displays the negative peaks exactly one octave above. As illustrated in supplementary material SEC. C, these negative peaks correspond to the frequency difference between partials in the case of odd-numbered harmonics. The negative peaks (and thus the frequency difference between partials) align with the pitches transcribed in Fig. 5's upper staff. Fig. 6(c) shows the distribution of frequencies corresponding to the resulting pitches. The intervals loosely conform to 12-TET, ranging from 20 to 40 ct above the 440 Hz tuning.

Fig. 7(a) shows the STFT for the same extract, zooming in on the lowest partial (f_0). The frequencies of this partial correspond to the pitches transcribed in Fig. 5's lower staff. Fig. 7(b) shows the distribution of pitches derived from the f_0. The intervals do not conform to 12-TET. Fig. 7(c) overlays the pitches derived from the f_0 (in color) with the pitches derived from autocorrelation (in white), transposed down one octave. The tuning of the autocorrelation-derived

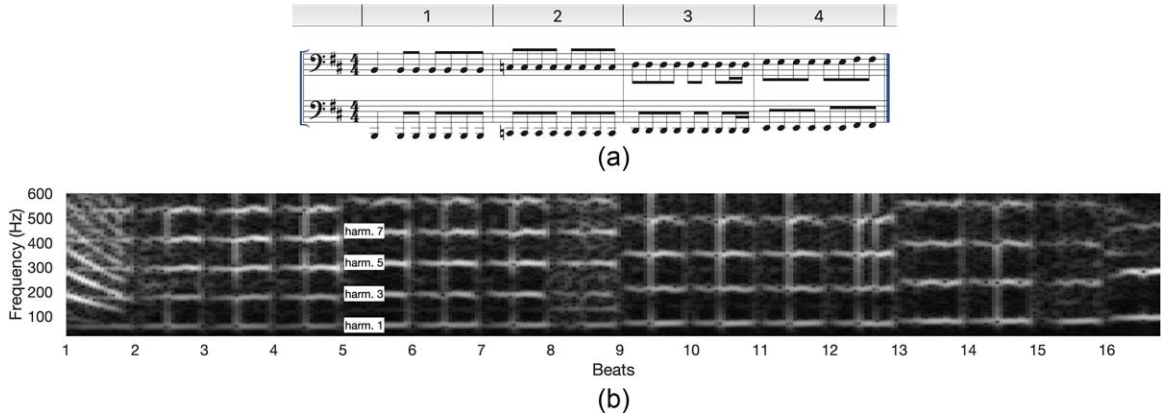


Fig. 5. "Next I'm Ready," 0'52 to 0'59. See supplementary material SEC. C for the same figure with synchronized audio. (a) Transcription of two perceived pitches. (b) STFT of the weighted audio. Even harmonics are weak or missing.

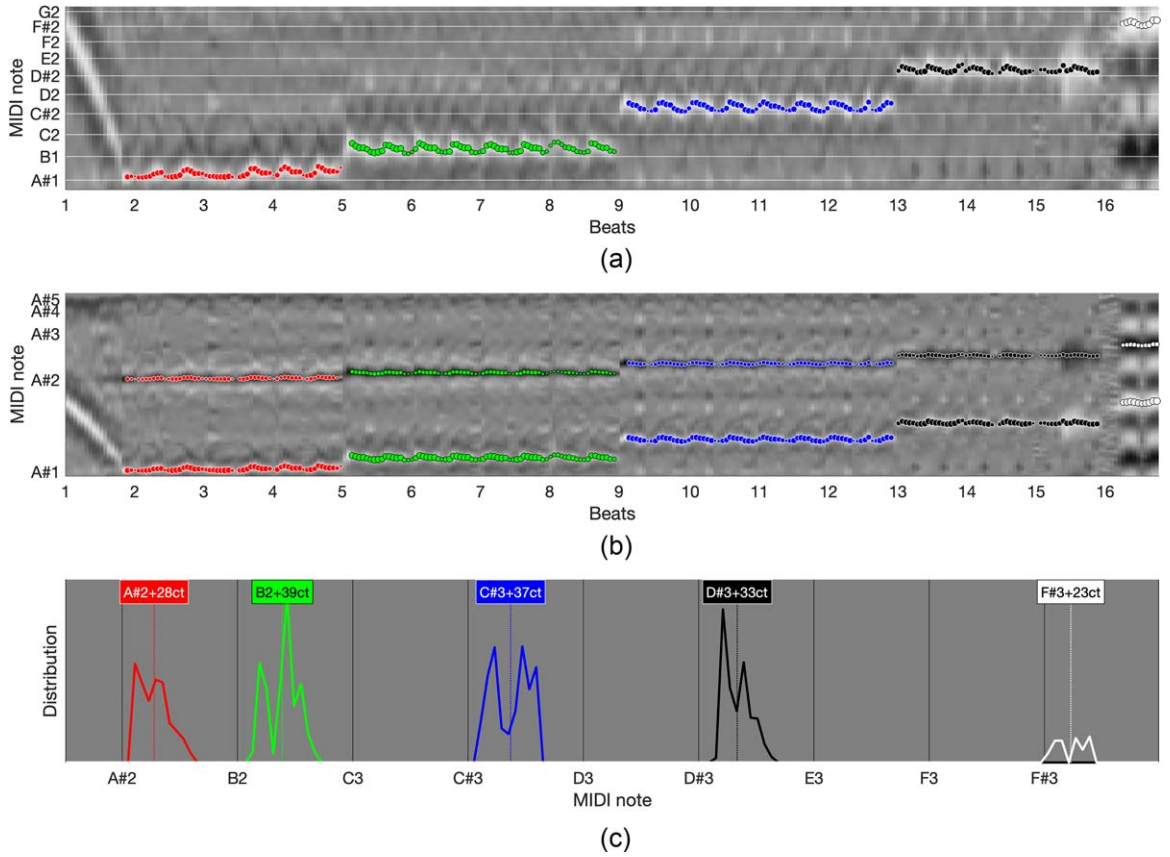


Fig. 6. "Next I'm Ready," 0'52 to 0'59. See supplementary material SEC. C for the same figure with synchronized audio. (a) STAC of the weighted audio, with the scatter plot showing maximum autocorrelation values. (b) STAC of the weighted audio, zoomed out so as to include higher frequencies. The upper and lower scatter plots show the STAC's local minima and maxima. If the y-scale is changed so that pitches distant from one octave are always equidistant, the two trajectories are parallel. (c) Distribution of pitches corresponding to the negative peaks, weighted by the corresponding correlation's absolute value. The labels show the weighted mean for each note.

pitches is generally lower than that of the f_0-derived pitches. Table 4 summarizes the differences between the pitches derived from autocorrelation and those from the f_0. There are variable differences between the tunings of each sequence.

2.2.3 Conclusion

The bass track can be described as two melodic lines stemming from a sequence of single complex tones (in the sense that each complex tone consists of a set of discrete partials that share "common fate"). The upper line's pitches are derived from temporal modeling (autocorrelation, frequency difference between partials), whereas the lower line's pitches are derived from spectral processing (f_0). The tuning corresponding to the f_0 does not comply with 12-TET and differs from that of the frequency difference between partials, which may explain the perception of the overall bass track as dissonant and out of tune. Vitalic confirms that the above properties are deliberately implemented, with such precise settings made possible by the use of software synthesizers—in this case, Native Instruments' Massive.

2.3 "Cosmic Renegade," Bass and Lead Synthesizer

2.3.1 Context and Motivation

"Cosmic Renegade" [28] is an instrumental, dance-oriented track. As shown in Fig. 8, it features a chord progression (B♭ min – B♭ min – A♭ maj – G♭ maj) with a repeating three-note synthesizer pattern over a bass part. The pattern sounds out of tune with the bass. In this analysis, the perceived pitches in both the bass and the pattern

Table 4. "Next I'm Ready," 0'52 to 0'59.*

	'Note' 1	'Note' 2	'Note' 3	'Note' 4	'Note' 5
Autocorrelation	A♯2 + 28 ct	B2 + 39 ct	C♯3 + 37 ct	D♯3 + 33 ct	F♯3 + 23 ct
F_0	A♯1 + 25 ct	C2 − 28 ct	D2 − 44 ct	E2 − 38 ct	F♯2 + 48 ct
Difference (cent)	3	33	19	29	25

* Comparison between the pitches corresponding to each note, autocorrelation [Fig. 5(a), upper staff] and f0 (lower staff).

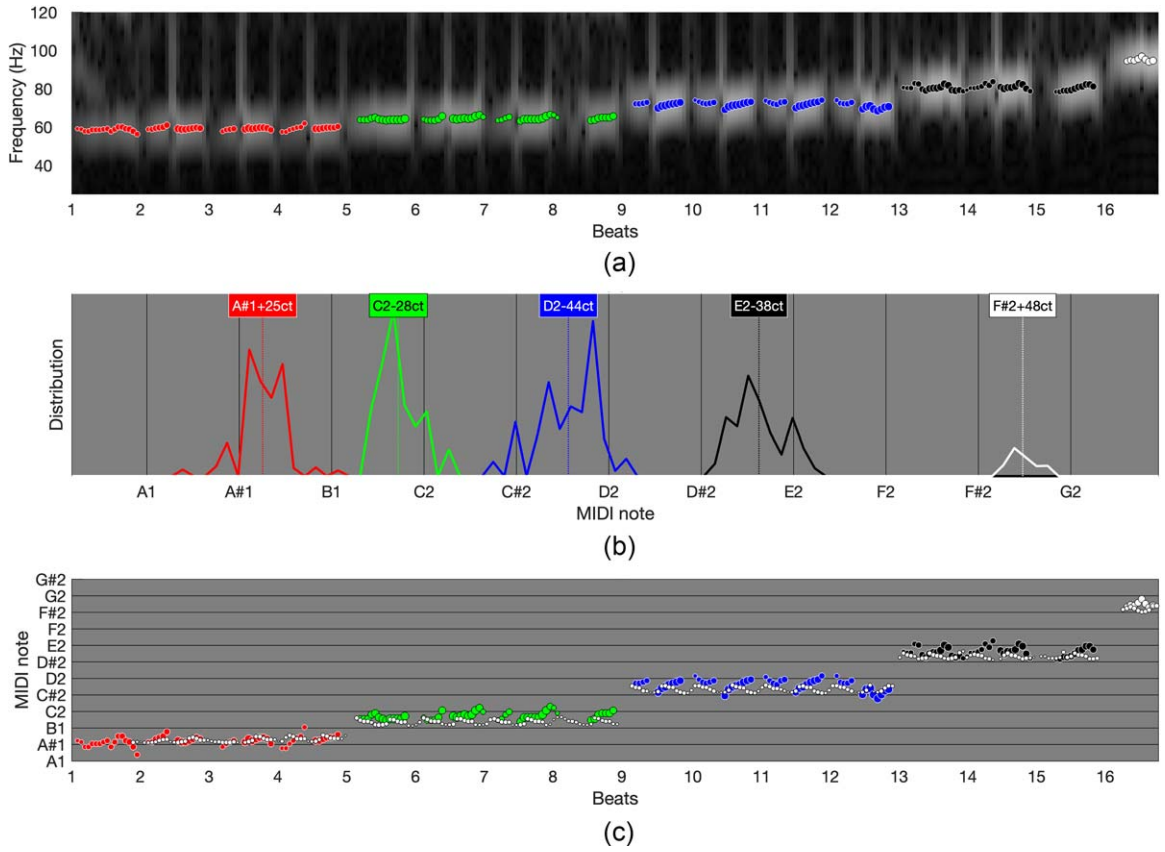


Fig. 7. "Next I'm Ready," 0'52 to 0'59. See supplementary material SEC. C for the same figure with synchronized audio. (a) STFT of the weighted audio showing the first partial (f_0). (b) Distribution of the frequencies corresponding to this partial, weighted by the partial's energy. The text boxes show the weighted mean for each note. (c) Comparison of the frequencies derived from the f_0 (in color) and those derived from autocorrelation (in white).

are linked to specific signal characteristics and explanations are offered for why the pattern may sound out of tune.

2.3.2 Analysis

Fig. 8 shows an approximate transcription of some perceivable pitches in "Cosmic Renegade," 2'55 to 3'02. The analysis focuses on the fourth bar in the extract, where the bass and upper parts are most distinct, making the spectrograms more readable than in the first three bars.

Fig. 9 focuses on the bass part. Fig. 9(a) shows that, whereas the notes in Fig. 8 (bottom staff) change, the part is based on a single complex tone with a fundamental near G♭1, although this fundamental is not directly perceived. The shifting pitches are linked to subsets of partials, similar to SEC. 2.1. The part was created using Native Instruments' Massive, with voltage-controlled filters enabling the amplification of different partials over time. Fig. 9(b) shows how the f_0 of the closest harmonic tones align with equal temperament, with the bass tuning 46 ct above 440 Hz. Figs. 9(c) and (d) illustrate how the perceived pitches align with autocorrelation peaks, resulting in similar estimations.

Fig. 10 focuses on the synthesizer part. Fig. 10(a) tracks the f_0 and overtones, leading to the f_0s of the closest harmonic tones shown in Fig. 10(b), which deviate slightly from the original f_0s. The synthesizer tuning deviates from equal temperament, with the B♭3–C4 interval being 15 ct smaller, likely a result of the deliberate use of the wow ("drift") effect from the Baby Audio Super VHS plug-in during production. More significantly, the synthesizer tuning is almost a quarter tone lower than the bass, confirming the initial perception of detuning.
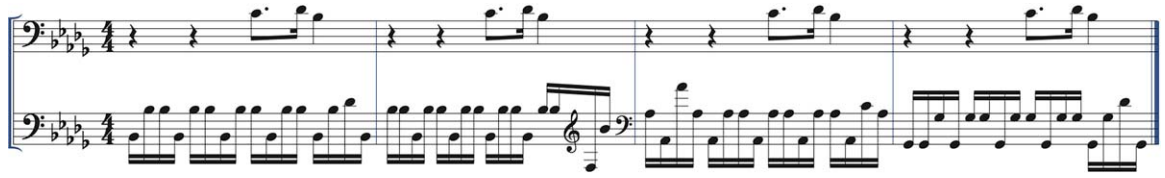


Fig. 8. "Cosmic renegade," 2'54 to 3'02, bass (bottom) and lead synthesizer (top), approximate transcription. See supplementary material SEC. D for the same score with synchronized audio.

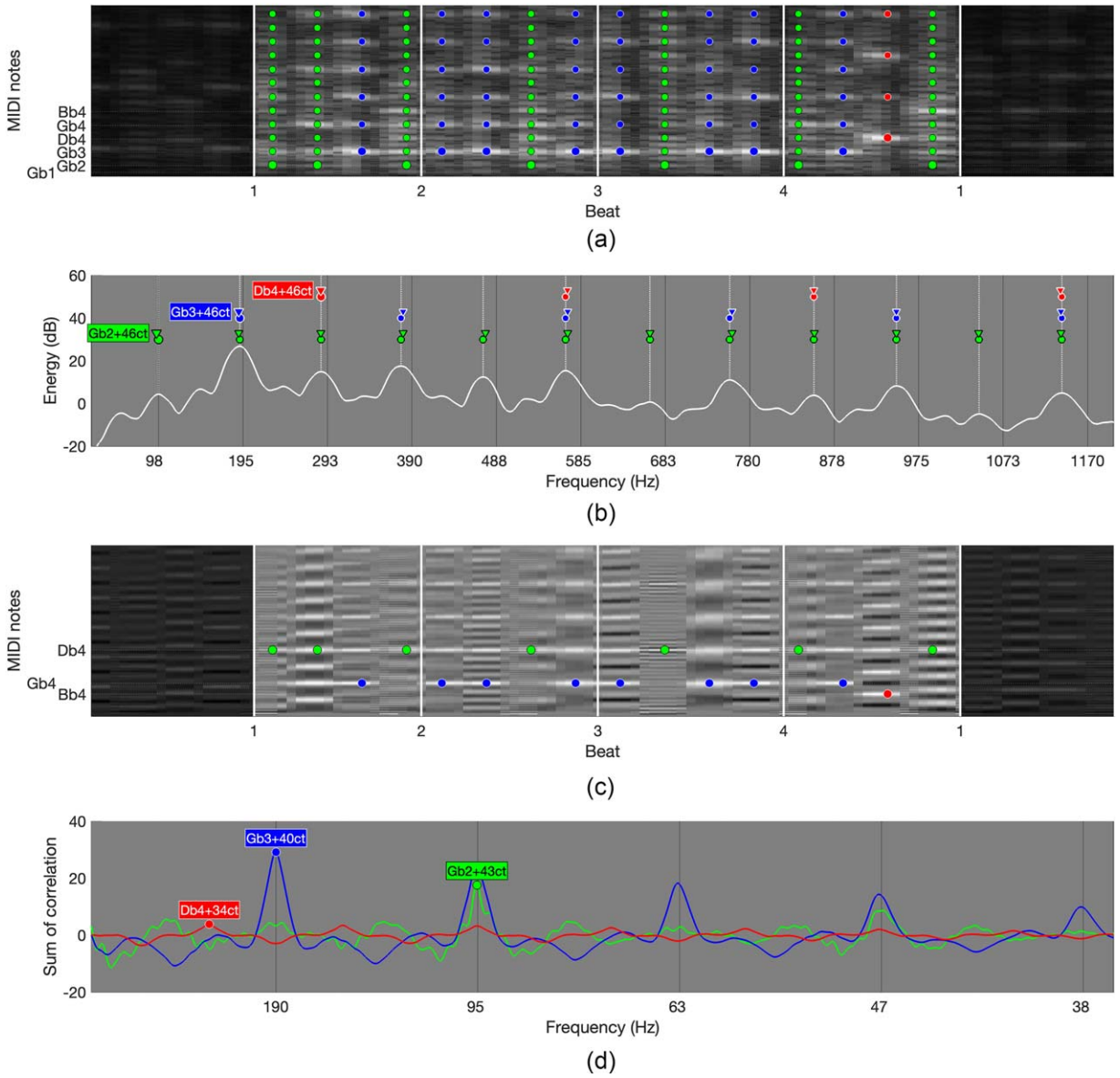


Fig. 9. "Cosmic Renegade," 2'54 to 3'02, bass. Corresponds to Fig. 8, bottom staff, fourth bar. See supplementary material SEC. D for the same figure with synchronized audio. (a) STFT with weighted audio. Scatter plots show formant positions and perceived pitches (green = G♭2, blue = G♭3, red = D♭4). (b) STFT for the entire extract. Circular points indicate formant positions, triangles indicate harmonic position of the closest harmonic tone. Labels give f_0 of the closest harmonic tone. (c) STAC with weighted audio. Dots show positions of perceived pitches. (d) Autocorrelation for the entire bar. Peaks correspond to the pitches transcribed in Fig. 8.

2.3.3 Conclusion

The bass part follows 12-TET, and each perceived pitch in this part corresponds to a subset of the partials present in the sound. Pitches in the lead synthesizer part primarily stem from the f_0 and deviate slightly from 12-TET. The bass and synthesizer tunings differ by a quarter tone, explaining why the synthesizer is perceived as out of tune.

2.4 "Levitation," Lead Synthesizer

2.4.1 Context and Motivation

"Levitation" [29] is an upbeat, mostly instrumental electronic track. A lead synthesizer part plays from 2'34 to 3'38, repeating every eight beats. The synthesizer pitches sound out of tune, making transcription to staff notation difficult and, arguably, inappropriate. In this analysis, the perceived pitches in the synthesizer pattern are linked to specific signal characteristics and confirm that the pitches strongly deviate from 12-TET.

2.4.2 Analysis

The lead synthesizer part was isolated using source separation. Fig. 11 summarizes the analysis. Fig. 11(a) shows the STFT of the weighted audio, highlighting the lead keyboard's partials. Fig. 11(b) displays the autocorrelation peaks for the same audio. Information from Fig. 11(a) and Fig. 11(b) is summarized in Fig. 11(c), which simultaneously shows the musical instrument digital interface (MIDI) pitches corresponding to the tone's lowest partial (yellow), the autocorrelation analysis (blue), and the closest harmonic tone's f_0 (red).

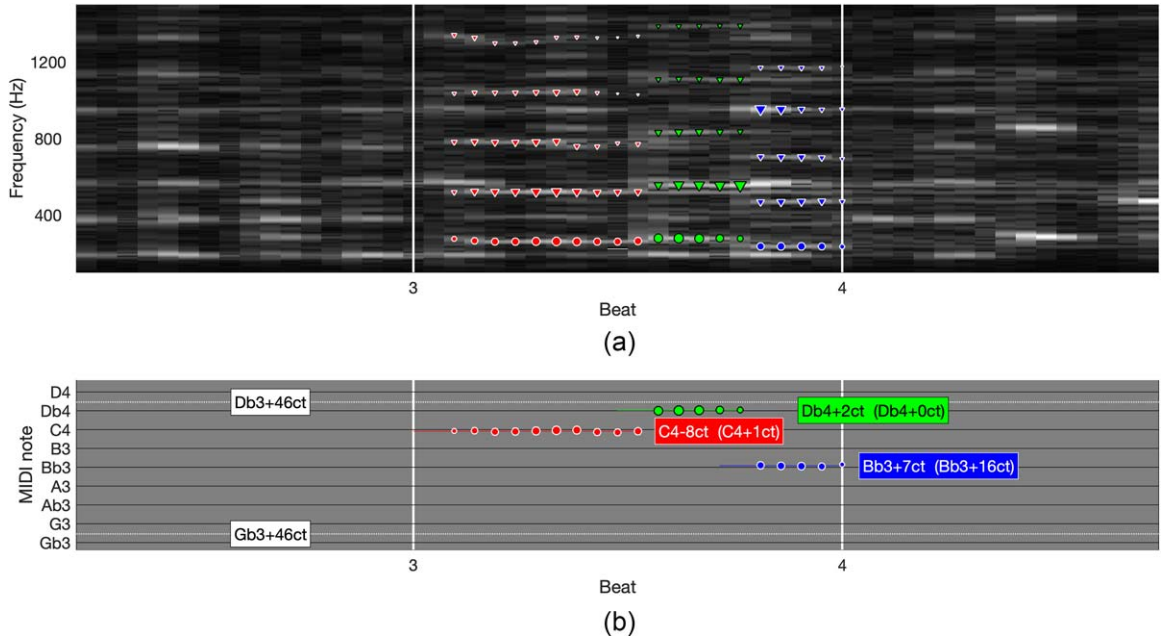


Fig. 10. "Cosmic Renegade," 3'00 to 3'02, lead synthesizer. Corresponds to fourth bar of top staff in Fig. 8. See supplementary material SEC. D for the same figure with synchronized audio. (a) STFT with weighted audio. Circular dots indicate lowest partials (f_0). Triangles indicate first four overtones. (b) f_0s of the closest harmonic tone (colored) compared with the bass pitches from Fig. 9 (indicated by white horizontal lines). The pitches in parentheses correspond to the original f_0s.

The frequencies within each note evolve continuously, with an ambitus that may reach up to one tone. The closest harmonic tone's f_0 mostly aligns with the autocorrelation peaks. The original f_0 values tend not to coincide with the autocorrelation peaks (i.e., the tone tends to be more inharmonic) when the note begins, ends, or during periods of significant pitch evolution. This behavior is reminiscent of transient states in acoustic instruments reported by Fletcher [6].

Fig. 11(d) shows the pitch distribution and mean value for each note. The mean values (B2 − 6 ct, D3 + 6 ct, E3 + 9 ct, G3 + 14 ct, A3 + 30 ct, C4 + 25 ct, C♯4 + 14 ct, and D4 + 33 ct) deviate significantly from 12-TET. It can be argued that they do not align with any specific tuning system because pitches from one pitch class assume two different values (D3 + 6 ct and D4 + 33 ct).

2.4.3 Conclusion

The lead synthesizer in "Levitation" follows a melodic line comprising a discrete sequence of elements with pitches that evolve continuously to some extent. Vitalic describes the sequence as being constructed not from notes but from distinct yet homogeneous "pitched objects." The perceived pitches are primarily derived from temporal modeling. The difficulty in transcribing the synthesizer part into a score arises from 1) unstable pitch contours and 2) pitch distributions whose peaks do not conform to 12-TET.

The part was created using XferRecords' Serum, a software wavetable synthesizer. The pitches originate from both Serum's oscillators and its reverberation, which is set to resonant settings, tuned to the oscillator's pitch, and modulated by the same low-frequency oscillator (LFO) (SEC. 2.6 provides another example of how reverberations contribute to pitch). Vitalic often finds that digital synthesizers produce a less rich output than their analogue counterparts, so he tends to detune the pitches to give the sound a more "vintage" and interesting character. The significant detuning is also justified by the context: a dance floor/disco setting that remains stylistically stable, allowing for more variations in the lead part.

2.5 "Polkamatic," Synthesizers

2.5.1 Context and Motivation

"Polkamatic" [30] opens the 2005 album *OK Cowboy*. Vitalic explains that the synth parts in this fast-paced electronic track were designed to evoke the sound of a slightly broken street organ gradually accelerating. The section from 0'00 to 0'12 features detuned synthesizer arpeggios and upward glissandi with a gradually increasing tempo. This analysis focuses on the perception of tones as being out of tune, which the artist confirms is central to the piece's effect.

2.5.2 Analysis

Fig. 12 summarizes the introduction to "Polkamatic." An arpeggio-based pattern is repeated three times, superimposed with desynchronizd, upward glissandi. The f_0 distribution for the arpeggio pattern does not comply with 12-TET. Fig. 13 tracks the first four partials in the initial pattern. The tones are inharmonic, with each tone (E♭5, G♭5, and B♭5) showing distinct distributions of distances between partials. The tuning of the tones is influenced by the distance between partials. The f_0 distribution for the closest equivalent harmonic tones differs from that of the f_0 of the original tones. It is close to 12-TET.

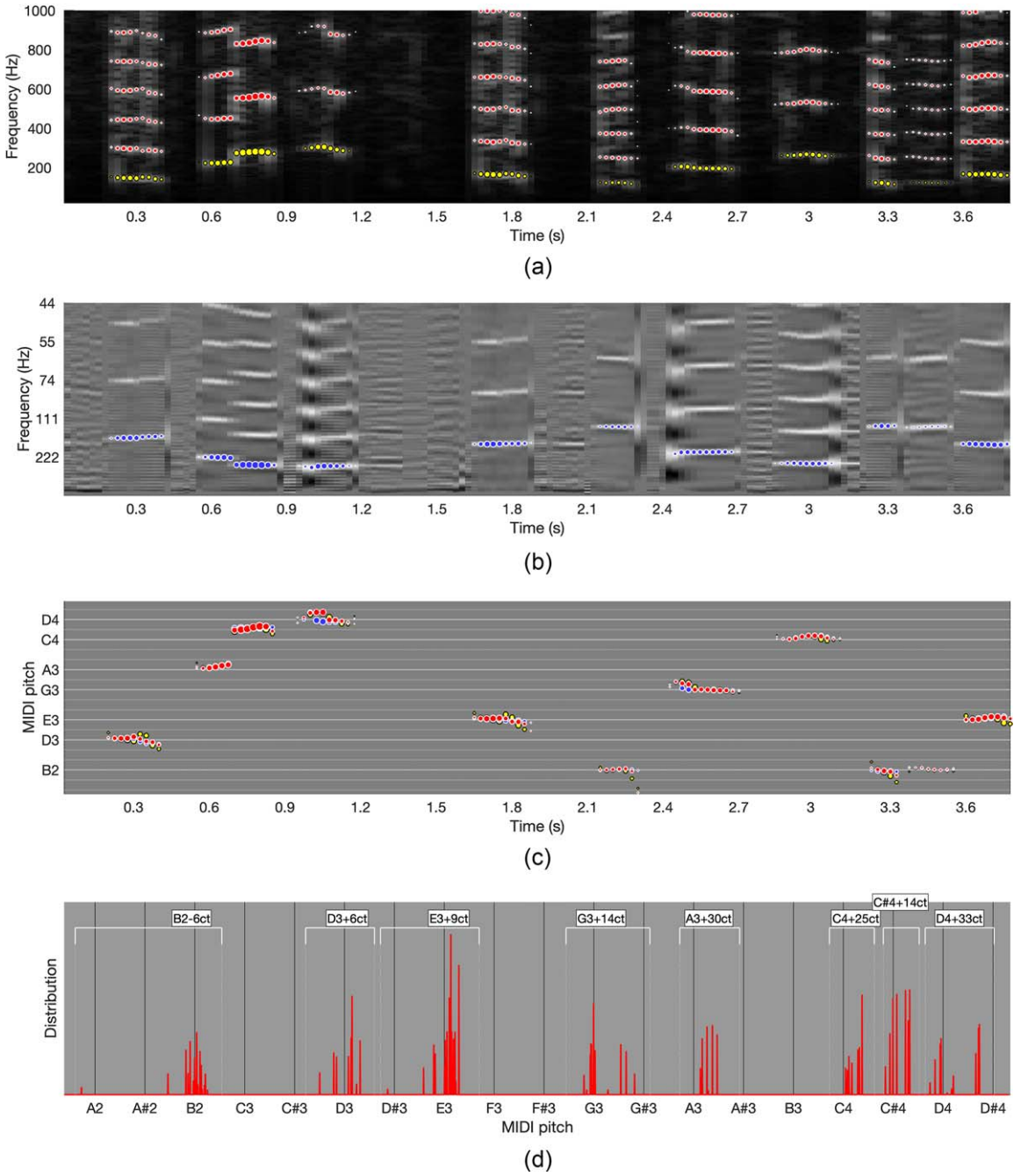


Fig. 11. "Levitation," 2'38 to 2'42, lead synthesizer part. See supplementary material SEC. E for the same figure with synchronized audio. (a) STFT of the weighted audio. Yellow dots track the lowest partial (f_0), with dot size representing energy (automatic tracking). Red dots track the first five overtones (tracked near the multiples of the f_0). (b) STAC. Blue dots (same size as yellow dots) mark the autocorrelation peak near the lowest partial in (a). (c) Comparison of frequency values from (a) (yellow), (b) (blue), and the closest harmonic tone's f_0 (red, same size as yellow). Yellow and blue dots are invisible when aligned with the harmonic tone's f_0. Here, the harmonic tone's f_0 mainly follows autocorrelation values. (d) Distribution for the harmonic tone's f_0, with mean value weighted by the lowest partial's energy.

A comparison of the distributions corresponding to the E♭5 tone (red) in Figs. 13(b) and (d) suggests that for this tone, instances where the original f_0 is lowest show a larger frequency difference between partials. This is confirmed by Fig. 14: the frequency difference between partials in the E♭5 tones follows an upward movement for each note simultaneously with the downward movement of the original f_0 seen in Fig. 12(a).

Video F5 in supplementary material SEC. F includes the resynthesis of the arpeggio part from 0'00 to 0'04 with four tunings: a) original, b) f_0s tuned to 12-TET, c) frequency differences between partials tuned to harmonic relations, and d) f_0s tuned to 12-TET with frequency differences tuned to harmonic relations. The results suggest that 1) the tuning of the f_0 has a greater impact than the tuning of partial frequency differences and 2) whereas the f_0 of the clos-

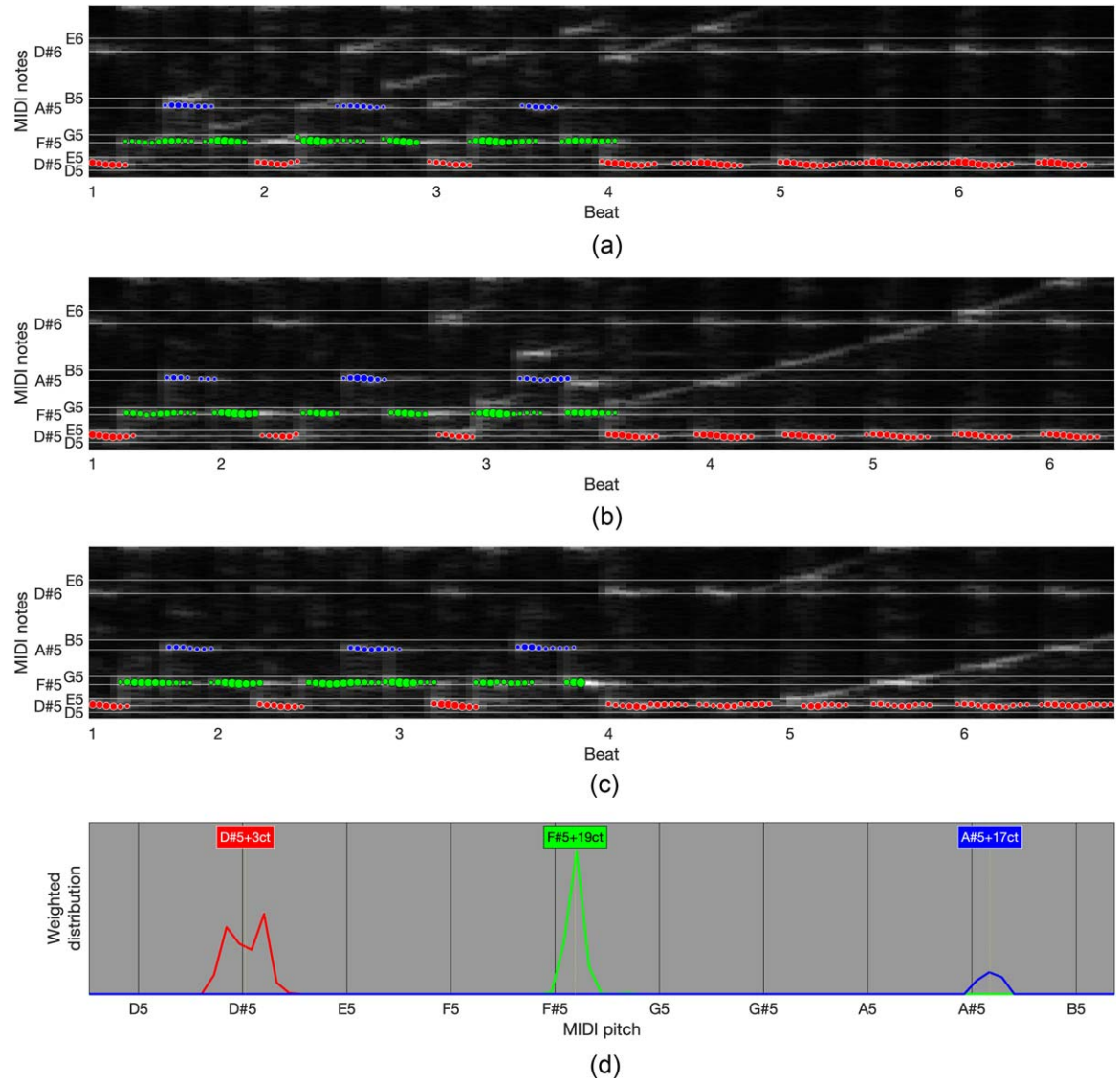


Fig. 12. "Polkamatic," 0'00 to 0'12. See supplementary material SEC. F for the same figure with synchronized audio. (a) "Polkamatic," 0'00 to 0'04, STFT of the weighted audio, with colored dots tracking the first partial of the keyboard part. Dot size represents energy at that frequency. (b) "Polkamatic," 0'04 to 0'08, STFT of the weighted audio. (c) "Polkamatic," 0'08 to 0'12, STFT of the weighted audio. (d) Distribution of the frequencies from (a), (b), and (c), weighted by energy of partials.

est harmonic tone is near 12-TET, the original tones in Fig. 12(a) sound more out of tune than the corrected tones Fig. 12(d). It may be concluded that the simultaneous contrary movements of the original f_0 and partial frequency differences create a sense of detuning while maintaining overall compliance with 12-TET.

Fig. 15 focuses on the glissandi, which use a tone similar to the arpeggios. The energy of the glissando between 0'00 and 0'04 is not constant; it is greater when crossing certain pitch values, which do not comply with 12-TET or coincide with the arpeggio pitches. This property of the glissando may derive from its implementation, which uses an LFO on both stable and unstable voltage-controlled oscillators in an Alesis A6 Andromeda.

2.5.3 Conclusion

"Polkamatic" was designed to resemble a broken street organ, featuring arpeggios and glissandi that emphasize certain pitches, both outside of 12-TET. The deviations are linked to a combination of f_0s and frequency differences between partials. The relationship between f_0s and the frequency differences is not random.

2.6 "Lightspeed," Lead Synthesizer

2.6.1 Context and Motivation

"Lightspeed" [31] is an instrumental, dance-oriented track with no clear chord progression. As shown in Fig. 16, it features a syncopated bass and a synthesizer hook [55] following an "AAAC" form, with "C" denoting a contrast in the sense of Bimbot et al. [56]. The hook appears to deviate from equal temperament. In bars 7–8 of the "C" pattern, the last note (D4) is played through reverberation, using Soundtoys Crystallizer with additional delays. Vitalic claims he frequently uses reverberations to generate pitch by increasing their resonance until they produce a defined

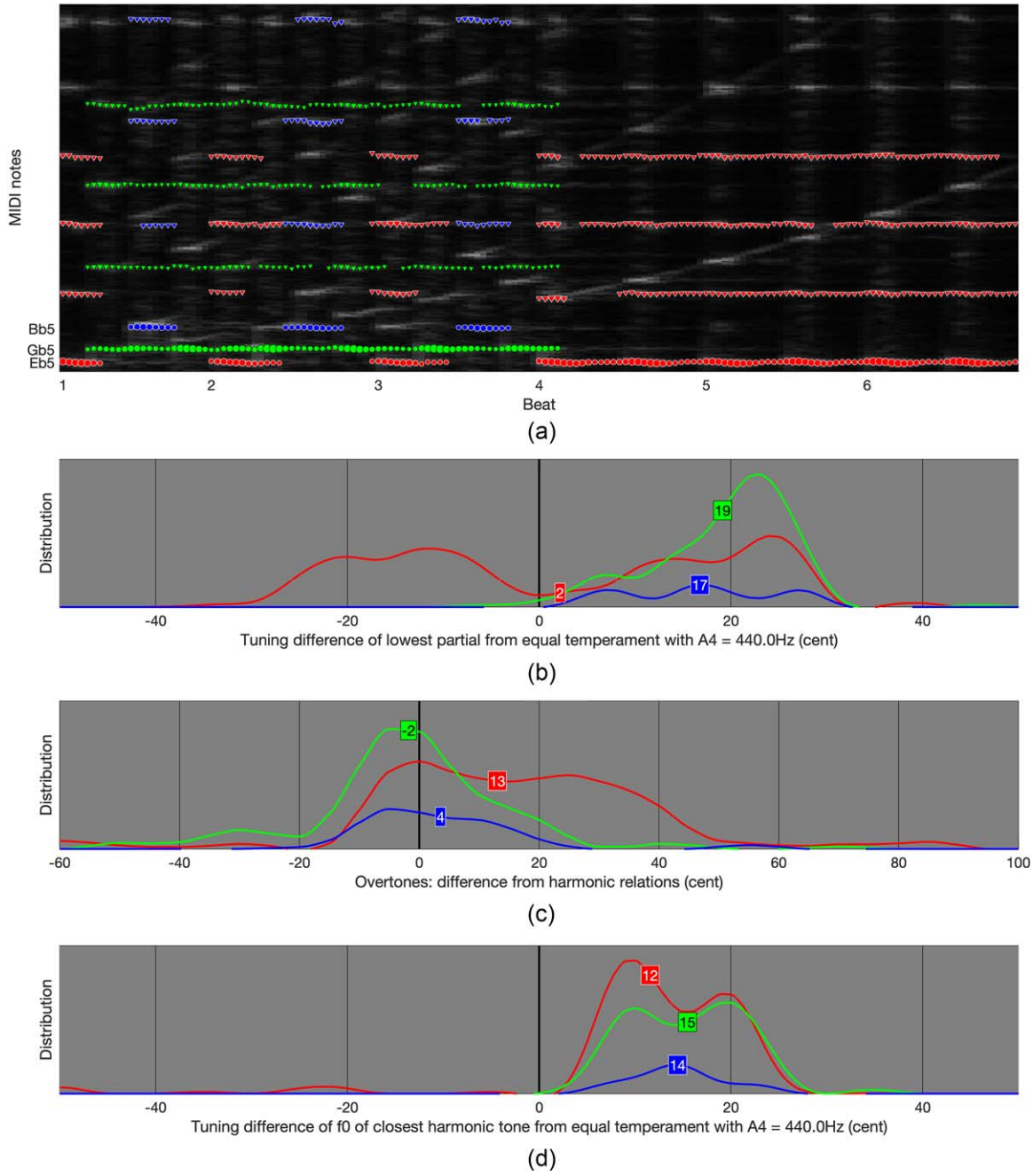


Fig. 13. "Polkamatic," 0'00 to 0'04. (a) STFT of the weighted audio. The circular dots represent the f_0 for each note: red for E♭5, yellow for G♭5, and blue for B♭5. The triangles represent the first four overtones. (b) Deviation of the lowest partial from equal temperament. The labels on the curves give the mean deviation for each note, weighted by the energy of the corresponding spectral peak. The colors are the same as in (a). (c) Deviations of the overtone positions from their harmonic positions. (d) Deviation from equal temperament of the f_0s of the closest harmonic tones.

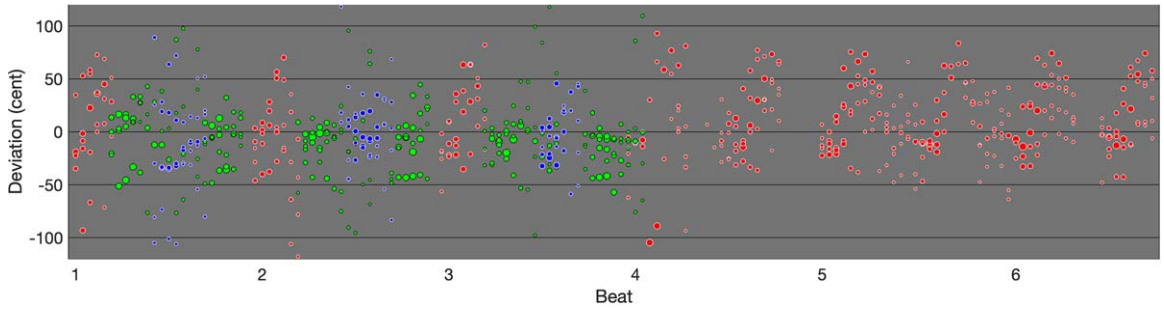


Fig. 14. "Polkamatic," 0'00 to 0'04. See supplementary material SEC. F for the same figure with synchronized audio. Deviation of the upper partials from equal temperament. Each point represents a partial. The colors refer to the different notes.

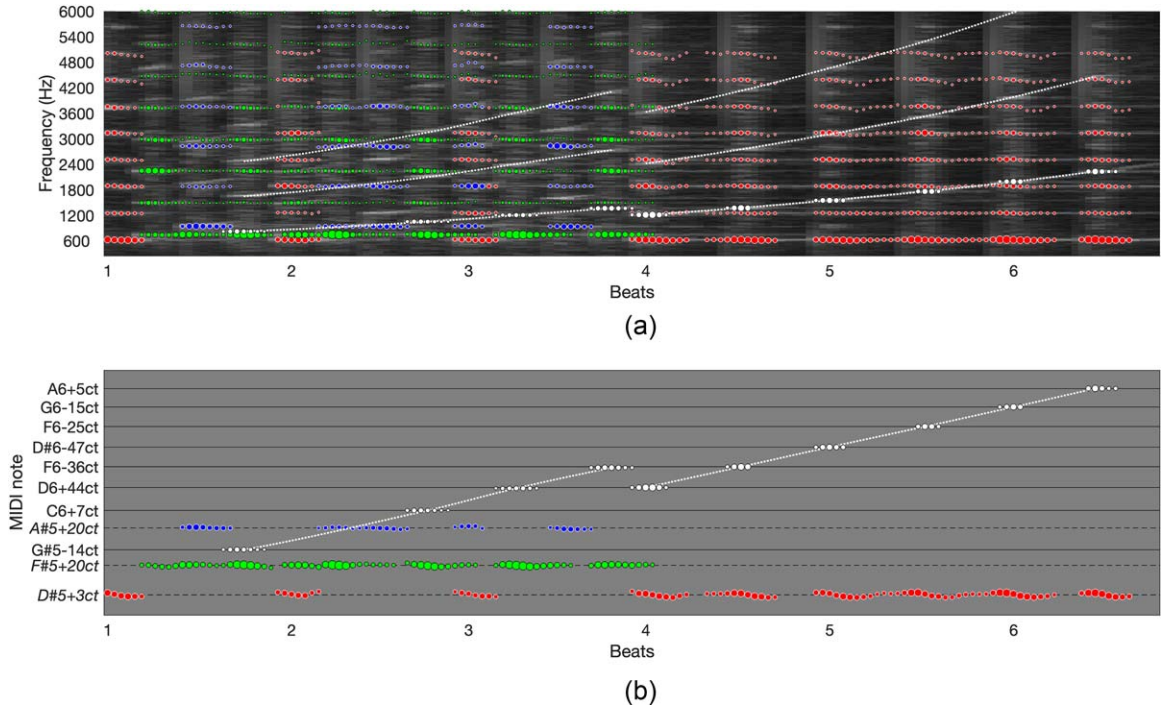


Fig. 15. "Polkamatic," 0'00 to 0'04. See supplementary material SEC. F for the same figure with synchronized audio. (a) STFT of the weighted audio. Colored dots track the partials of the keyboard part, and white dots track the glissando partials, with their size exaggerated for clarity. The thin dotted line represents the bicubic approximation of the glissando frequencies. (b) Representation of the lowest partials for both the keyboard and glissando parts. The thin dotted line again shows the bicubic approximation of the glissando frequencies. The italicized labels on the *y*-axis indicate keyboard notes, and regular labels indicate glissando notes.

pitch. These reverberations may be integrated within the synthesizer (see SEC. 2.4) or applied as separate plug-ins. The tuning of the reverberation's mode may override that of the original source. The analysis focuses on examining the reverberated part in terms of its deviation from 12-TET.

2.6.2 Analysis

Fig. 17(a) shows the STFT of the synthesizer part obtained through source separation, with tracked partials. The source separation algorithm retains some bass elements in the output.

There is significant variability in the partials' energies. For instance, the first partial (f_0) is heavily attenuated at the start of beat 2, and several partials disappear at the start of beat 3. Vitalic attributes this variability to the use of a sequencer and LFOs that individually control the partials, made possible with the Buchla Music Easel used to generate the pattern. The sequencer is programmed on a five-note cyclic pattern to avoid synchronization with the meter. In beats 7–8, the f_0 disappears entirely, this time owing to the reverberation.

This variability suggests that temporal modeling plays a shifting role throughout the extract. Pitch perception in bar 5 may involve spectral processing of the f_0, unlike in bars 7–8. The marker hues in Fig. 17(b) reflect this phenomenon, with darker hues indicating stronger f_0 influence.

Fig. 17(b) also confirms that the synth pattern does not comply with 12-TET. The extract can be divided into four sections: in beats 1–4, tuning rises within a quarter-tone range; in beat 5, tuning is lower, based on the f_0; in beat 6, incoming reverberation disturbs the tuning; and in beats 7–8, the reverberation, whose pitch is modulated independently of the source, causes the tuning to drop significantly.

2.6.3 Conclusion

The synthesizer pattern in bars 7–8 deviates from 12-TET because of the combined effects of temporal and spectral processing of f_0. The reverberation in beats 7–8 increases

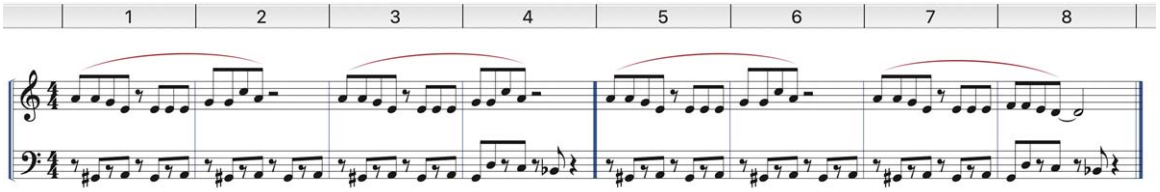


Fig. 16. "Lightspeed," 0'55 to 1'10, bass (bottom) and synthesizer (top), approximate transcription. See supplementary material SEC. G for the same figure with synchronized audio.

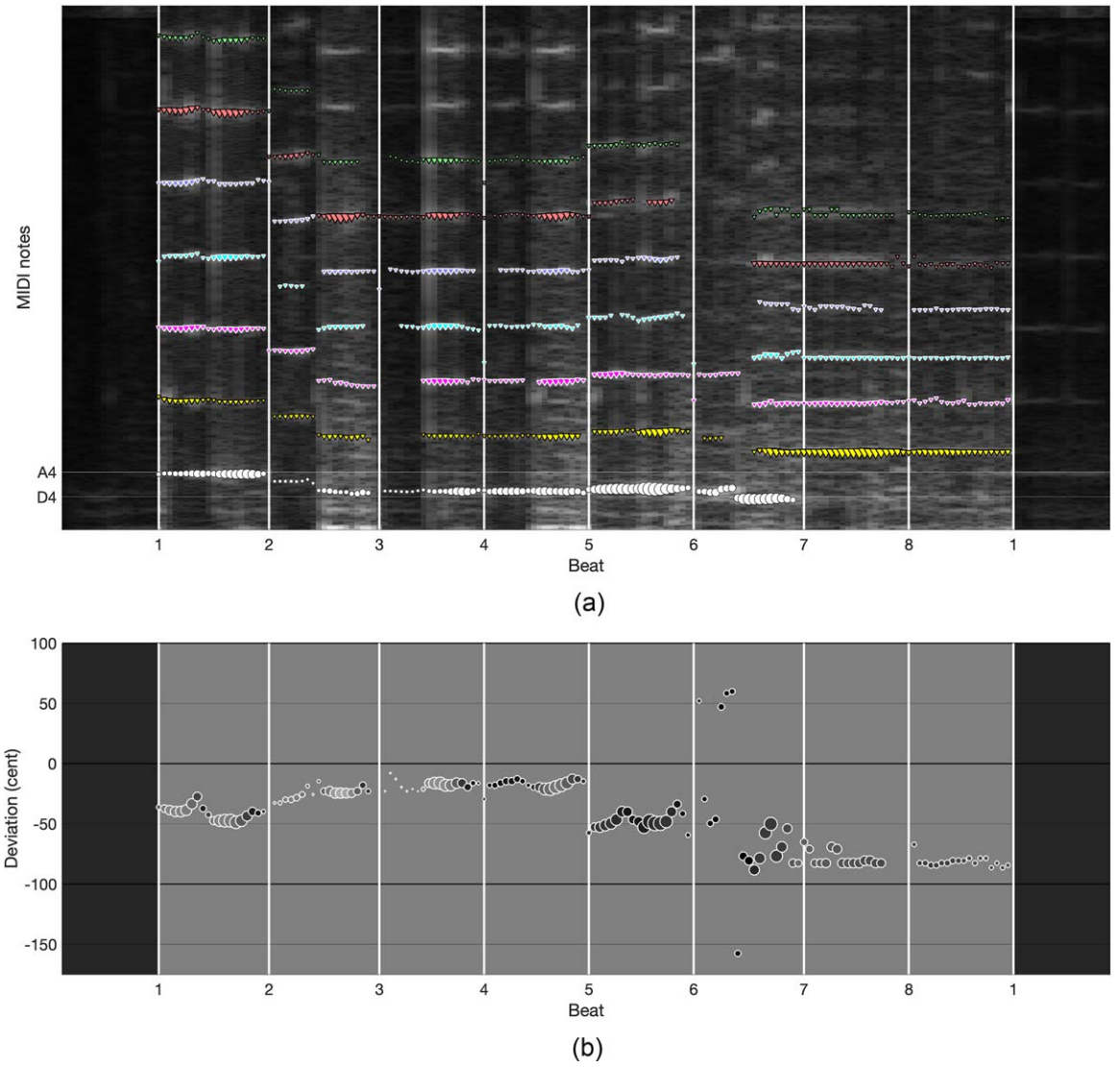


Fig. 17. "Lightspeed," 1'05 to 1'10, keyboard corresponding to bars 7 and 8 in Fig. 16. See supplementary material SEC. G for the same figure with synchronized audio. (a) STFT with nonweighted audio. Circular dots represent f_0s; triangles represent overtones. (b) Deviation from equal temperament (A4 = 440 Hz) of the f_0s of the closest harmonic tone. Marker size indicates overall energy. Darker markers indicate more energy in lower partials.

the influence of temporal modeling and further detunes the part. As in SEC. 2.4, the significant detuning is justified by the dance floor/disco context, which remains stylistically stable, allowing for more variation in the lead part.

2.7 "Winter Is Coming," Synthesizers

2.7.1 Context and Motivation

"Winter is Coming" [32] is the final piece on *Dissidænce Episode 2*. It features only synthesizers without drums. The section between 0'40 and 1'20 contains a stable synthesizer playing a chord sequence while another plays short, six-'note' elements on top. This analysis is motivated by the perception of some pitches in both tracks as out of tune. It links the perceived pitches in the six-'note' pattern and the chord to specific signal characteristics.

2.7.2 Analysis

Fig. 18 examines the chord and its relation to the six-note pattern, and Fig. 19 details the six-note pattern. Fig. 18(a) shows an approximate transcription of the 2.6-s extract. The six-'note' pattern is transcribed on the upper staff; the chord is shown on the lower staff.

Fig. 18(b) presents the STFT of the signal with highlighted partials. The STFT shows that the "chord" consists of partials, not superimposed complex tones. The lowest partial corresponds to A1, and the fifth partial is a natural C4 instead of a C♯4, giving the chord its minor aspect. Each partial's frequency evolves independently. The chord is created using Arturia's plug-in implementation of the Solina string ensemble with LFO pitch modulation, the Baby Audio Super VHS plug-in, and distortion.

The frequency of the six-note pattern's lowest partial (f_0) is unstable, with the ambitus reaching one half-tone for

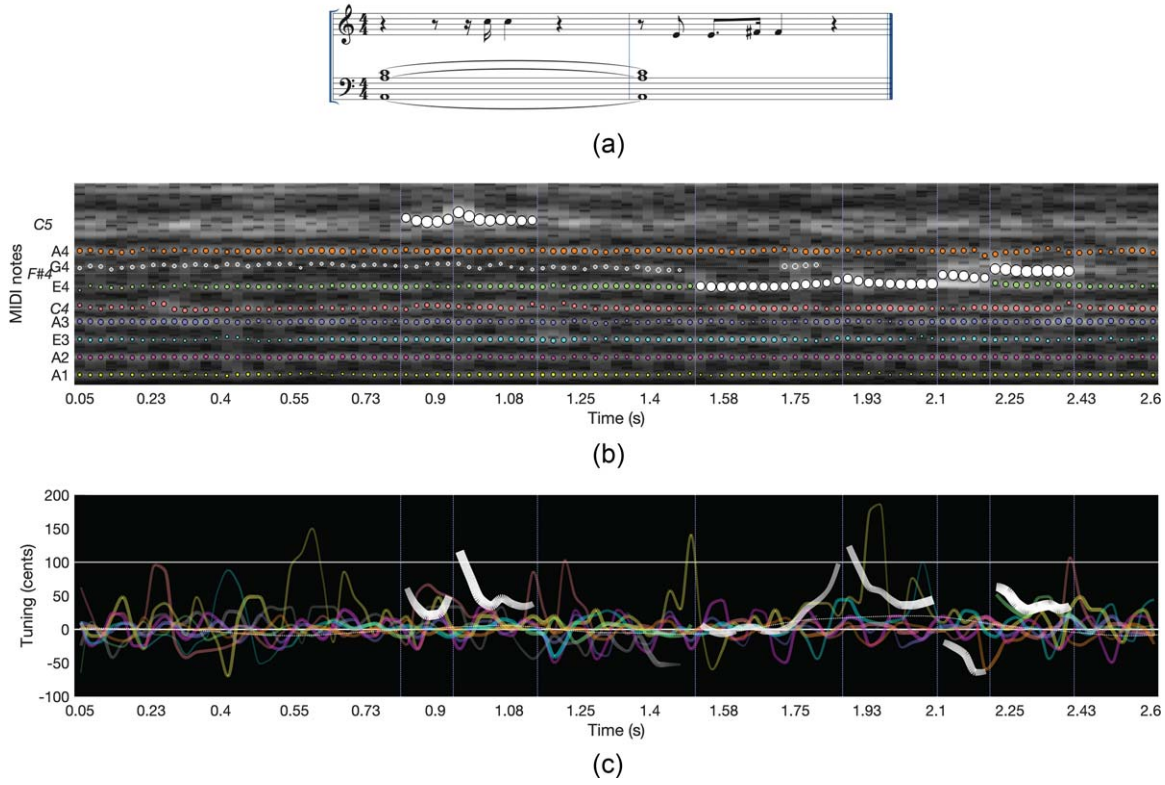


Fig. 18. "Winter Is Coming," 0'47 to 0'50. See supplementary material SEC. H for the same figure with synchronized audio. (a) Approximate transcription. (b) STFT of the weighted audio. Partials corresponding to the bottom staff in (a) are marked with colored dots, and those for the top staff are highlighted in white. The dot size represents the partial's energy, with white dots enlarged for clarity. On the y-axis, notes in upright font are harmonics of A1, and italicized notes are not. Notes to the left of others do not correspond to partials. (c) Deviation of frequencies of partials from 12-TET with A4 = 440 Hz. Colors match (b).

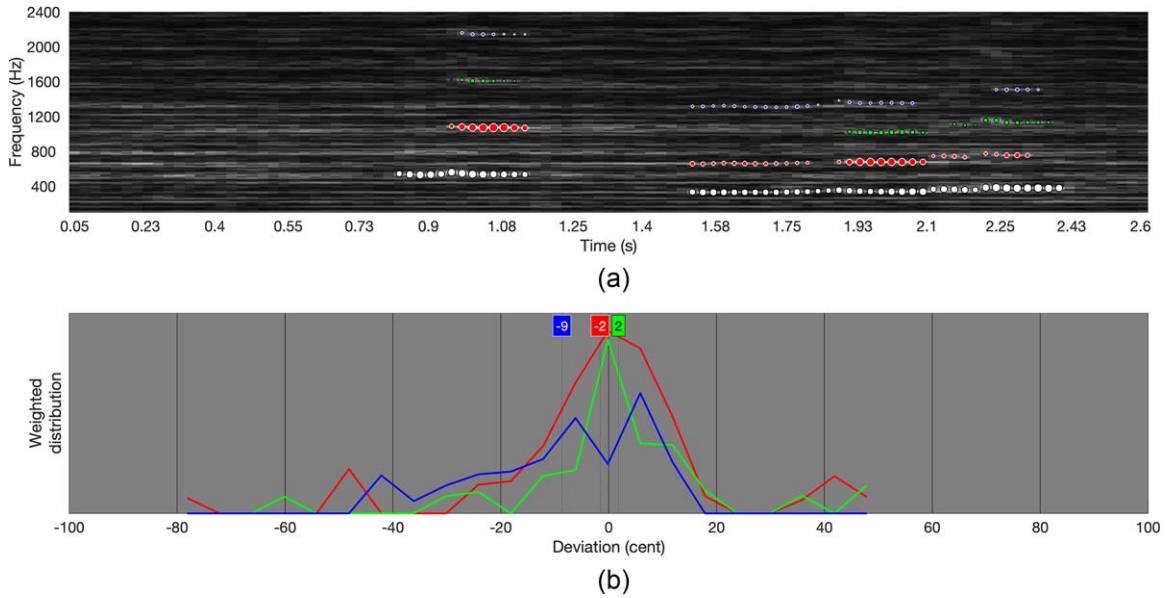


Fig. 19. "Winter Is Coming," 0'47 to 0'50. See supplementary material SEC. H for the same figure with synchronized audio. (a) First four partials corresponding to the top staff in Fig. 18. Partials are omitted when tracking fails. (b) Deviations of partials from harmonic positions.

Table 5. "Winter Is Coming," 0'47 to 0'50.*

Partial	1	2	3	4	5	6	7	8
Deviation (cent)	+20	0	+3	0	+8 (−92)	+3	0	−2
Theoretical deviation (cent)	0	0	+2	0	−14	+2	−31	0

* Second row shows deviation of partial frequencies from equal temperament (A4 = 440 Hz). Partial 5 is approximately one semitone lower than the equal temperament quantization of the theoretical harmonic position. Third row shows theoretical deviation of harmonic partials from 12-TET.

each note. Fig. 19 shows that the pattern uses a quasiharmonic complex tone, whose perceived pitch can, therefore, be linked to its f_0 value.

Fig. 18(c) and Table 5 summarize the tuning for each partial in the chord and the pattern's f_0. The pattern's f_0 tuning differs considerably from that of the chord and does not follow 12-TET. The chord partials are not tuned to frequencies harmonically related to the fundamental: partials 2–8 are tuned around 12-TET, with partial 1 being 20 ct higher.

2.7.3 Conclusion

The extract features a melodic pattern and a perceived chord. The pattern uses a quasiharmonic tone whose f_0 does not follow 12-TET, which may explain why it is perceived as being out of tune. The chord consists of partials, not simultaneous tones, with a fifth partial one semitone below the harmonic position. The upper partials' frequencies are centered on 12-TET instead of following harmonic relations, whereas the first partial is higher, which explains why it is heard as being out of tune.

2.8 Recapitulation

Table 6 summarizes the analyses provided in SECS. 2.1 to 2.7. All tones are inharmonic. Single tones may result in one pitch (various f_0 and temporal modeling contributions), two pitches (conflict between f_0 and temporal modeling, individually audible partials, audible partial subsets), or several pitches (individually audible partials). Deviation from 12-TET may result from the f_0 of the closest harmonic tone, the individually audible partials, or the audible partial subsets not being aligned to 12-TET.

Table 6. Recapitulation: number of simultaneous tones involved, number of simultaneous pitches involved, origin of pitches, and origin of the deviation from equal temperament.

Section	2.1 "No Fun," Bass	2.2 "Next I'm Ready," Bass	2.3 "Cosmic Renegade," Bass and Lead Synthesizer
Number of simultaneous tones involved	One sequence of single tones	One sequence of single tones	Two simultaneous sequences of single tones
Number of simultaneous pitches involved	Two	Two	Two
Origin of pitch(es)	Each top pitch: individually audible upper partial; each bottom pitch: partial subset	Each top pitch: temporal modeling; each bottom pitch: individually audible first partial (f_0)	Each top pitch: f_0 and temporal modeling; each bottom pitch: partial subset
Origin of the deviation from 12-TET	Each top pitch: partial not aligned with 12-TET; each bottom pitch: frequency difference between partials not aligned with 12-TET	Each top pitch: frequency difference between partials not aligned with 12-TET; each bottom pitch: partial frequency not aligned with 12-TET.	Each top pitch: f_0 of closest harmonic tone not aligned with 12-TET; each bottom pitch: frequency difference between partials not aligned with 12-TET

2.4 "Levitation," Lead Synthesizer	2.5 "Polkamatic," Synthesizers	2.6 "Lightspeed," Lead Synthesizer	2.7 "Winter Is Coming," Synthesizers
One sequence of single tones	Two simultaneous sequences of single tones	One sequence of single tones	One static tone, one sequence of single tones
One	Two	One	Static tone, several; single tone, one
Each pitch: f_0 and temporal modeling	Each pitch in each sequence: f_0 and temporal modeling	Each pitch: f_0 and temporal modeling (variable contributions)	Static tone: individually audible partials; each pitch in the sequence: f_0 and temporal modeling
Each pitch: f_0 of closest harmonic tones not aligned to 12-TET	Each pitch in each sequence: f_0 of closest harmonic tones not aligned to 12-TET	Each pitch in each sequence: f_0 of closest harmonic tones not aligned to 12-TET; further detuning by reverberation	Static tone: partials not in harmonic relation; each pitch in the sequence: f_0 of closest harmonic tones not aligned to 12-TET

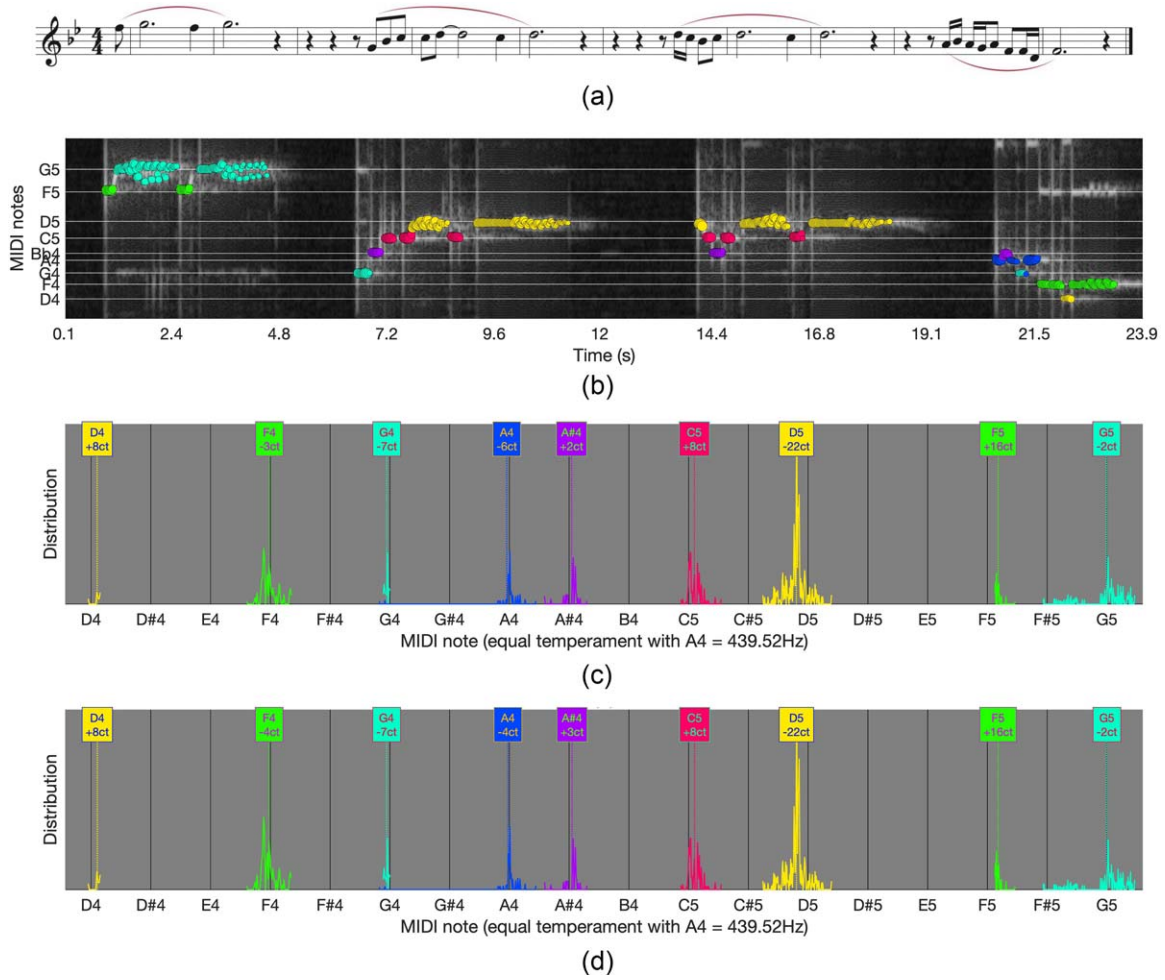


Fig. 20. Pink Floyd, "Shine On You Crazy Diamond," 2'10 to 2'34. See supplementary material SEC. I for the same figure with synchronized audio. (a) Approximate transcription. (b) STFT of the weighted audio. Scatter plots correspond to the f_0, with each perceived note in its own color. The 439.52-Hz value is derived from the simultaneous keyboard part. (c) f_0 distribution. Text boxes indicate the mean f_0 for each note, weighted by the frame's energy. (d) Distribution of f_0s of closest harmonic tones.

3 EXAMPLES FROM ARTISTS OTHER THAN VITALIC

This section presents three examples of pitch sequences deviating from 12-TET in music by artists other than Vitalic, demonstrating that this practice is not unique to Vitalic's work. SEC. 3.1 showcases a highly expressive electric guitar performance. SEC. 3.2 focuses on an inharmonic synthesizer part. Finally, SEC. 3.3 illustrates how commercially available software easily provides access to overtones that deviate from 12-TET.

3.1 Pink Floyd, "Shine on You Crazy Diamond," Guitar

Part 1 of "Shine On You Crazy Diamond" [37] features a 1-min, 30-sec guitar solo, with pitch deviating from 12-TET. Fig. 20(a) shows an approximate transcription of the first 24 s. Fig. 20(b) tracks the evolution of the lowest partial (f_0) from the guitar output of the source separation process. Fig. 20(c) presents the frequency distribution by note. Although the guitar tone is inharmonic (see supplementary material Figure I2), Fig. 20(c) suggests that accounting for inharmonicity (f_0 of the closest harmonic tone; see supplementary material Fig. I3) does not significantly alter the overall distribution.

A key feature is the flattened D5, the most prominent note, whose tuning creates a narrower fifth in G minor. As in SEC. 2.4, these deviations do not establish a consistent alternative tuning system because different octaves of the same pitch class are tuned differently.

Deviation from 12-TET in the "Shine On You Crazy Diamond" guitar solo results at least partly from vibrato and string-bending, a technique used by guitarists to explore microtonality and continuous pitch variation [57]. The effects of vibrato and string-bending are visible in Fig. 20(b): the initial transition from F5 to G5 results from string-bending, whereas the oscillating movement below the subsequent G5 is produced by vibrato.

Because the deviation arises from manual instrumental gestures, it may vary with each performance, as seen in live versions such as [58]. This contrasts with live interpretations of Vitalic, where the prerecorded tone sequence is often played back as in "No Fun," "Lightspeed," and "Levitation" [59].

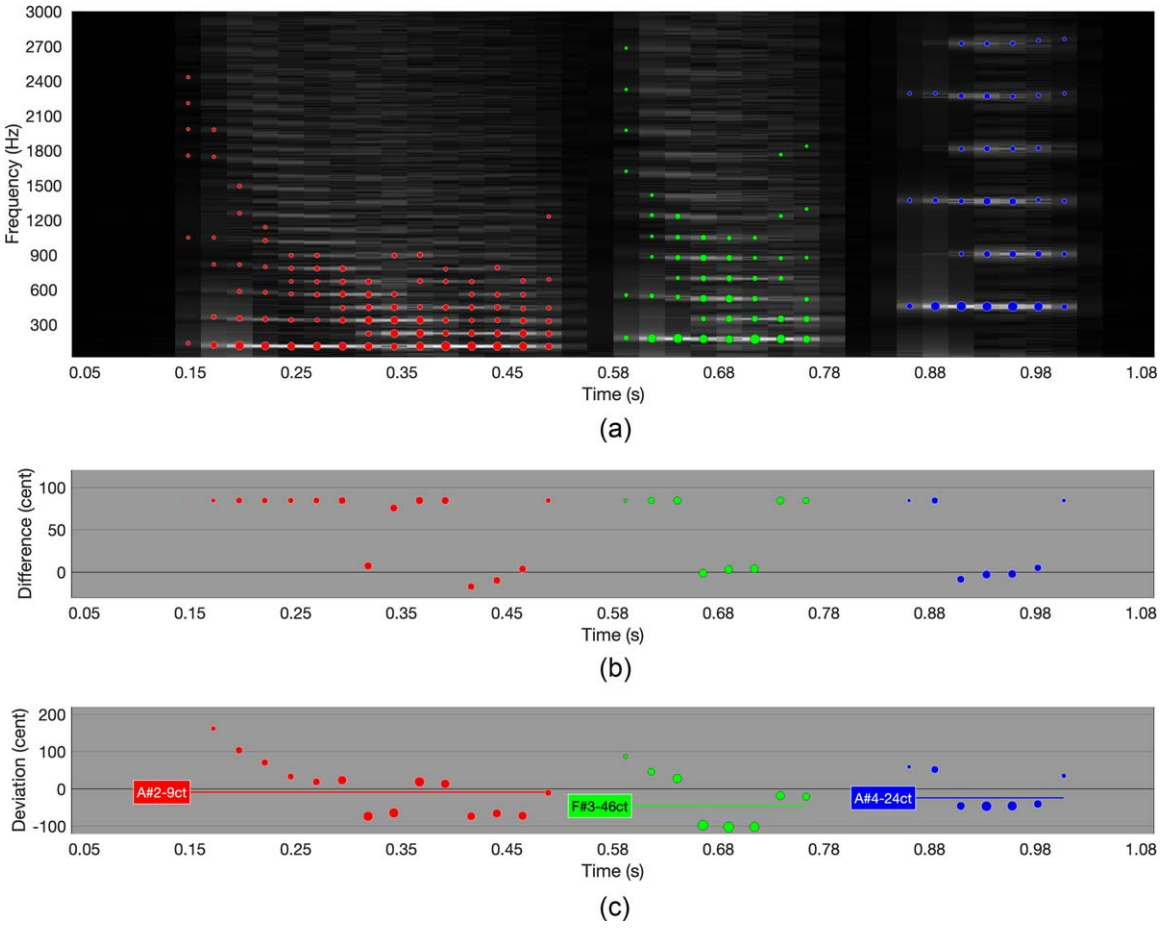


Fig. 21. Kanye West, "On Sight," synthesizer, 2'01 to 2'03. See supplementary material SEC. J for the same figure with synchronized audio. (a) STFT of the weighted audio. Colored scatter plots track the first six partials. Each color corresponds to a note. (b) Difference between the f_0 of the original tone and the f_0 of the closest harmonic tone. (c) Deviation from 12-TET of the f_0 of the closest harmonic tone. The horizontal colored lines show the weighted mean for each note.

3.2 Kanye West, "On Sight," Synthesizer

In "On Sight" [38], a song produced with Daft Punk, the most prominent pitched element, apart from the vocals, is a rhythmic synthesizer part that repeats and alters a brief pattern throughout the song. Fig. 21(a) shows the STFT of one occurrence of the pattern near the 2-min mark. Fig. 21(b) emphasizes the inharmonicity of the tone by illustrating the difference between its f_0 and the nearest harmonic tone's f_0. Fig. 21(c) demonstrates the deviation from 12-TET, with the second note of the pattern tuned 37 ct lower than the first. Other instances of the pattern show varying frequency values (see supplementary material video J1).

3.3 Primaal, "Whomp," Hook

Primaal is a brand from Hyper Music,[1] a sound and music production company that creates music for advertisements, television series, and feature films. This example, taken from Primaal's music, relates to SECS. 2.1 and 2.7, where perceived pitches that do not comply with 12-TET derive from individually audible overtones. Multiple pitch perception from a single tone in music by the same production company has previously been documented [60].

"Whomp" includes a hook track, a simple melody using a sample from an unspecified source. The sample is processed with Sony Computer Science Laboratories Resonance EQ [61], boosting the partials. Two patterns are audible simultaneously: "G♭2 F2 F2 F2 F0" and "B♭4 A4 A4 A4 A4." The interval between them is two octaves plus a major third, aligning with the producers' aim to make harmonic 5 audible. Fig. 22 (b) suggests that the perception of A4 originates from a subset of F2's overtones, recalling the process used in "No Fun" (SEC. 2.1).

4 DISCUSSION

4.1 A (More) General Framework for Musical Pitch

Given a tone as defined in SEC. 1.2, this study identifies three types of correspondence between the signal and perceived pitches:

1. A quasiharmonic tone prompts the perception of a single pitch. An approximation of this pitch may

[1] https://www.hyper-music.com/

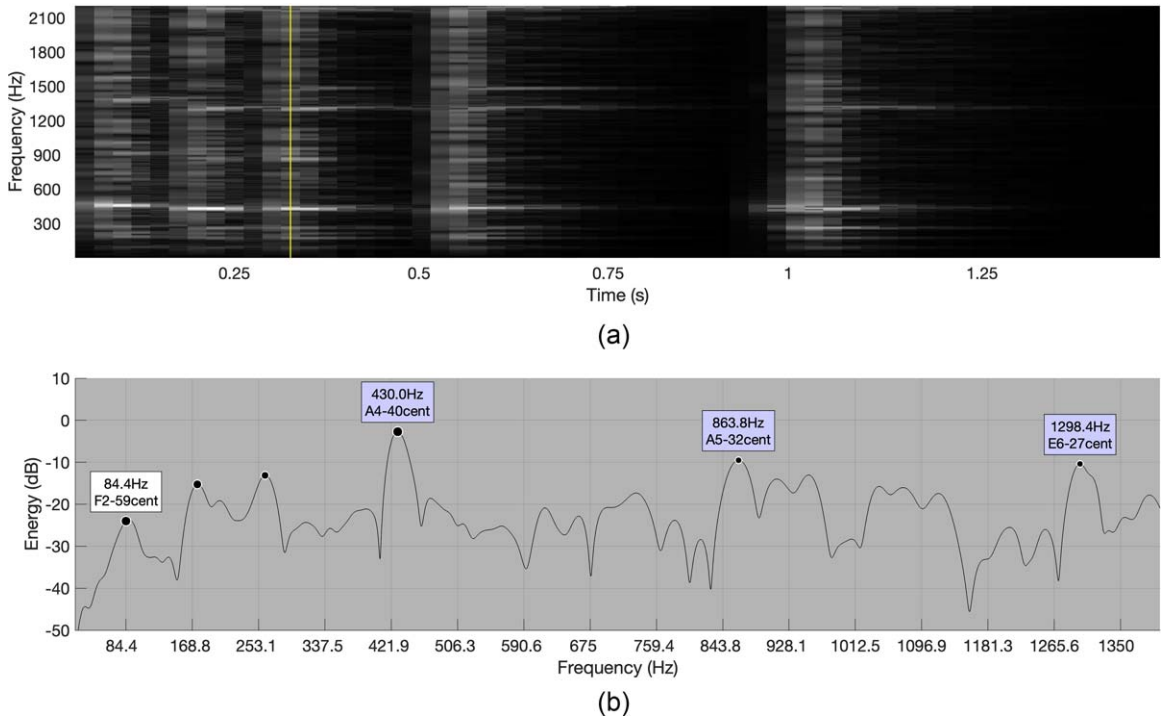


Fig. 22. "Whomp," hook track, 0'48 to 1'05. See supplementary material SEC. K for the same figure with synchronized audio. (a) STFT with weighted audio. (b) Fourier transform at frame indicated by vertical yellow line in (a). The grid shows the theoretical positions of the harmonics of the lowest partial. The blue backgrounds indicate the partials corresponding to the upper audible pitch.

be the f_0 of the closest harmonic tone. If the tone is harmonic, this value corresponds to the lowest partial.

2. A tone prompts the perception of multiple pitches derived from the positions of its partials. Mistuned partials may be perceived as separate pitches, as documented by Moore et al. [43]. In some cases, the mistuned partial itself can correspond to the perceived f_0, as in "Next I'm Ready" (SEC. 2.2). Temporal modelling may conflict with spectral modelling in such cases.
3. A tone prompts the perception of multiple pitches derived from the energy distribution of its partials. Subsets of overtones may be perceived as distinct pitches. This phenomenon relates to the human ability to "hear out" individual harmonics. First introduced by Mersenne [62], this concept became central to the Ohm-Seebeck dispute, where Ohm argued that only the fundamental partial determines the perceived pitch of a tone [63]. Experiments suggest humans can typically hear the first 5–8 harmonics, although their audibility remains contested [64]. In Vitalic's music, overtones are intentionally made audible.

The traditional framework of acoustic instruments corresponds to situation 1 above. The tone is considered quasi-harmonic [65], resulting in a single pitch [66]. In this context, the frequencies and positions of partials remain within bounds: no partial is sufficiently loud, in the sense of Plomp [67], or mistuned, in the sense of Moore et al. [43], to be heard independently. Within these bounds, pitch can be separated from timbre.

In the general case involving situations 2 and 3, such bounds are not observed. Partials may be sufficiently loud or mistuned to be perceived independently. The pitches perceived depend on the individual amplitudes and positions of the partials. Consequently, pitch cannot be separated from timbre, resulting in a pitch-timbre continuum.

4.2 Active and Resultant Pitch

In Vitalic's compositional approach, the tone often comes first, with pitch positions selected from the available partials. In several examples from this study—most notably Vitalic's "No Fun," "Next I'm Ready," "Winter is Coming," "Cosmic Renegade" (bass), and Primaal's "Whomp"—the pitch values emerge as a consequence of the audio synthesis process.

The transition from modal polyphonic writing to harmonic and tonal writing has been described as involving a shift from resultant harmony to active harmony [68–70]. A parallel can be drawn between pitch resulting from the tones and harmony resulting from the polyphony in Renaissance counterpoint. Contemporary popular music with nonacoustic tones may involve a transition from active pitch to resultant pitch. In this framework, the pitches are resultant, and the tones are active.

4.3 The Musical Note as Carrier of Pitch Information

The transcription (encoding) of pitched content as Western scores stands on the following conventions. The written

pitches correspond to the 12-TET degrees that are closest to the audible pitches. The transcription will then be interpreted (decoded) by assigning a tone to each written pitch, and expressive variations will be added. Most examples in the study may be approximately transcribed to a Western score, with finer pitch information being discarded as "expressivity," although studies on expressivity do not include pitch [71].

Still, to what extent does the transcription convey the relevant information? In some cases, such as the arpeggio in "Polkamatic" (SEC. 2.5) and the guitar in "Shine On You Crazy Diamond" (SEC. 3.1), the transcription may be accurately reconstructed if the notes are complemented with pitch expressivity. This expressivity can be minimized by specifying a different temperament.

In other cases, the encoding/decoding scheme is simply unsuitable for representing pitch. For example, in "Winter is Coming" (SEC. 2.7), a key aspect of the musical discourse is the lowering of the fifth harmonic so that it is a compound minor third above the fundamental. Here, the fine value of the pitch is critically important, and an encoding that discards this information would fail to properly describe the music.

Perhaps even more importantly, accurately encoding or describing "Winter is Coming" would require specifying the frequency of each partial. However, by convention, encoding these frequencies as notes on a score implies that each frequency will be sonified using a complex tone, resulting in music that is fundamentally different. This issue, rooted in the pitch-timbre continuum, applies to all parts where a single tone results in the perception of more than one pitch.

In "Levitation" (SEC. 2.4), the measure used here for pitch varies continuously within a one-tone ambitus, which is higher than the 12-tone scale's resolution. Vitalic himself does not refer to these "objects" as "notes," but rather as "homogeneous pitched objects." The weighted mean of the pitch measure for each object does not adhere to 12-TET. In this case, representing the pitched content as a score is not meaningful. An f_0 trajectory would be more adequate.

Outside Vitalic's music, the synthesizer part in Kanye West's "On Sight" (SEC. 3.2) repeats throughout much of the song. The information conveyed by a transcription into musical notes is extremely sparse. Conversely, the information conveyed by expressivity varies with each repetition. In this case, the actual message may not lie in the notes, but in the expressivity itself. In light of the examples shown in this study, in general, within contemporary popular music, the musical note alone cannot be considered as carrying the relevant pitch information, and in some cases, the musical note is simply not the right encoding.

4.4 Keyboard Decorrelation

In Vitalic's music, the deviations observed from 12-TET may be significant enough for him to refer to a decorrelation that can occur between the keyboard notes and the audible pitches. In an informal interview, one of Primaal's producers, Luc Leroy, states that in such cases, the keyboard keys do not correspond to defined pitches. Instead, they are generic MIDI controllers, with lower pitches generally on the left and higher pitches on the right (although not always).

5 CONCLUSION

The original aim of this study was to relate signal features to pitch in the works of electronic music producer Vitalic to substantiate the perception that these pitches do not conform to 12-TET. To achieve this goal, it was necessary to identify the following relationships between perceived pitches and the signal that are not widely documented in the literature:

1. In quasiharmonic tones, pitch may be associated with the f_0 of the closest harmonic tone or with the f_0 itself if the tone is harmonic.
2. In the case of mistuned partials within an otherwise quasiharmonic context, both the context's pitch and the mistuned partials may be perceived.
3. Even in harmonic tones, subsets of overtones that are sufficiently loud can be perceived as individual pitches.

The results were as follows:

1. None of the examples in this study conform to 12-TET.
2. In some examples, pitches might be represented using a 12-semitone tuning system with nonequal temperament.
3. In other examples, pitches are better described as an f_0 trajectory.
4. In yet other examples, the accurate transmission of the pitched content requires specifying the partials' positions and amplitudes.

Ultimately, this study challenges the traditional view of musical pitch. We hope it offers new perspectives for analyzing and generating music, by reconsidering the role of pitch in both technical and creative contexts.

6 ACKNOWLEDGMENT

The authors would like to thank Yann Macé and Luc Leroy from the music production company Hyper Music for their advice and assistance. They also wish to thank Michael Turbot from Sony A.I. for his support.

7 REFERENCES

[1] J. Serrà, Á. Corral, M. Boguñá, M. Haro, J. L. Arcos, "Measuring the Evolution of Contemporary Western Popular Music," *Sci. Rep.*, vol. 2, no. 1, pp. 1–6 (2012 Jul.). https://doi.org/10.1038/srep00521.

[2] M. Mauch, R. M. MacCallum, M. Levy, A. M. Leroi, "The Evolution of Popular Music: USA 1960–2010," *R. Soc. Open Sci.*, vol. 2, no. 5, paper 150081 (2015 May). https://doi.org/10.1098/rsos.150081.

[3] E. Benetos, S. Dixon, Z. Duan, S. Ewert, "Automatic Music Transcription: An Overview," *IEEE Signal Process. Mag.*, vol. 36, no. 1, pp. 20–30 (2018 Dec.). https://doi.org/10.1109/MSP.2018.2869928.

[4] J. W. Kim, J. Salamon, P. Li, J. P. Bello, "Crepe: A Convolutional Representation for Pitch Estimation," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 161–165 (Calgary, Canada) (2018 Apr.). https://doi.org/10.1109/ICASSP.2018.8461329.

[5] A. Riou, S. Lattner, G. Hadjeres, G. Peeters, "Pesto: Pitch Estimation With Self-Supervised Transposition-Equivariant Objective," in *Proceedings of the 24th International Society for Music Information Retrieval Conference*, pp. 535–544 (Milan, Italy) (2023 Nov.). https://hal.science/hal-04260042/.

[6] N. H. Fletcher, "The Nonlinear Physics of Musical Instruments," *Rep. Prog. Phys.*, vol. 62, no. 5, p. 723 (1999 May). https://doi.org/10.1088/0034-4885/62/5/202.

[7] W. A. Yost, "Pitch Perception," *Atten. Percep. Psycho.*, vol. 71, no. 8, pp. 1701–1715 (2009 Nov.). https://doi.org/10.3758/APP.71.8.1701.

[8] B. Haynes, P. Cooke, "Pitch," *Grove Music Online* (2001). https://doi.org/10.1093/gmo/9781561592630.article.40883 (accessed April 10, 2025).

[9] M. Lindley, "Equal Temperament (Fr. *tempérament egal*; Ger. *gleichschwebende Temperatur*; It. *temperamento equabile*)," *Grove Music Online* (2001). https://doi.org/10.1093/gmo/9781561592630.article.08900 (accessed April 10, 2025).

[10] J. M. Barbour, *Tuning and Temperament: A Historical Survey* (Courier Corporation, North Chelmsford, MA, 2004).

[11] D. Morton, C. Duriyanga, *The Traditional Music of Thailand*, vol. 8 (University of California Press, Berkley, CA, 1976).

[12] J. Fonville, "Ben Johnston's Extended Just Intonation: A Guide for Interpreters," *Perspect. New Music*, vol., 29, no. 2 (summer), pp. 106–137 (1991 Aug.). https://doi.org/10.2307/833435.

[13] S. Cao, *A Parametric Analysis of Measures 0-51 Of Pithoprakta by Iannis Xenakis and an Original Composition Làng Táo Shā*, Ph.D. thesis, East Carolina University, Greenville, NC (2020 Jul.). https://www.proquest.com/openview/0affb5864a4a4f8b1657868cc091e33e/1.

[14] K. Choi, J. Park, W. Heo, S. Jeon, J. Park, "Chord Conditioned Melody Generation With Transformer Based Decoders," *IEEE Access*, vol. 9, pp. 42071–42080 (2021 Mar.). https://doi.org/10.1109/ACCESS.2021.3065831.

[15] E. Deruty, M. Grachten, S. Lattner, J. Nistal, C. Aouameur, "On the Development and Practice of AI Technology for Contemporary Popular Music Production," *Trans. Int. Soc. Music Inf. Retr.*, vol. 5, no. 1, pp. 35–39 (2022 Feb.). https://doi.org/10.5334/tismir.100.

[16] Diplo feat. Nicky Da B, "Express Yourself," (2012 Mar.) on *Express Yourself*, Mad Decent Protocol, EP. https://youtu.be/BKaL7WL-onI.

[17] E. Deruty, L. Leroy, Y. Macé, D. Meredith, "Methods for Pitch Analysis in Contemporary Popular Music: Highlighting Pitch Uncertainty in Primaal's Commercial Works," *arXiv preprint arXiv:*2502.08131 (2025 Feb.).

[18] H. Phares, "Vitalic Biography," Allmusic.com (2024). https://www.allmusic.com/artist/vitalic-mn0000220024#biography (accessed April 10, 2025).

[19] Vitalic, "And it Goes Like," on *Confess EP,* Citizen Records/Virgin Music France, EP (2023 Oct.).

[20] Vitalic, "Eternity," on *Voyager*, Clivage Music/Universal (2017 Jan.).

[21] Vitalic, "Nozomi," on *Voyager*, Clivage Music/Universal (2017 Jan.).

[22] J.-M. Jarre, "Équinoxe," Dreyfus/Sony Music (1978 Aug.).

[23] Sparks, "Tryouts for the Human Race," on *N.1 in Heaven,* Elektra-Virgin (1979 Mar.).

[24] MusicRadar Tech, "Vitalic In The Studio With Future Music" (2013). https://youtu.be/ogjeXBY-iEI (accessed April 10, 2025).

[25] E. Zwicker, H. Fastl, "Pitch and Pitch Strength," in *Psychoacoustics: Facts and Models*, Springer Series in Information Sciences, pp. 111–147 (Springer-Verlag, New York, NY, 1990). https://doi.org/10.1007/978-3-662-09562-1_5.

[26] Vitalic, "No Fun," on *OK Cowboy*, Different Records/PIAS (2005 Apr.).

[27] Vitalic, "Next I'm Ready," on *Rave Age*, Different Records/PIAS (2012 Nov.).

[28] Vitalic, "Cosmic Renegade," on *Dissidænce (Episode 1)*, Clivage Music (2021 Oct.).

[29] Vitalic, "Levitation," on *Voyager*, Clivage Music/Universal (2017 Jan.).

[30] Vitalic, "Polkamatic," on *OK Cowboy*, Different Records/PIAS (2005 Apr.).

[31] Vitalic, "Lightspeed," on *Voyager*, Clivage Music/Universal (2017 Jan.).

[32] Vitalic, "Winter is Coming," on *Dissidænce (Episode 2)*, Clivage Music (2022 Feb.).

[33] R. J. Ritsma, "Existence Region of the Tonal Residue. I," *J. Acoust. Soc. Am.*, vol. 34, no. 9A, pp. 1224–1229 (1962 Sep.). https://doi.org/10.1121/1.1918307.

[34] R. Plomp, "Pitch of Complex Tones," *J. Acoust. Soc. Am.*, vol. 41, no. 6, pp. 1526–1533 (1967 Jun.). https://doi.org/10.1121/1.1910515.

[35] R. D. Patterson, F. L. Wightman, "Residue Pitch as a Function of Component Spacing," *J. Acoust. Soc. Am.*, vol. 59, no. 6, pp. 1450–1459 (1976 Jun.). https://doi.org/10.1121/1.381034.

[36] B. C. Moore, B. R. Glasberg, R. W. Peters, "Relative Dominance of Individual Partials in Determining the Pitch of Complex Tones," *J. Acoust. Soc. Am.*, vol. 77, no. 5, pp. 1853–1860 (1985 May). https://doi.org/10.1121/1.391936.

[37] Pink Floyd, "Shine On You Crazy Diamond," on *Wish You Were Here*, Harvest, Columbia Records (1975 Sep.).

[38] Kanye West, "On Sight," on *Yeezus*, Def Jam Recordings, Roc-A-Fella Records (2013 Jun.).

[39] Primaal, "Whomp," (2023). Alter-K Global Music Services.

[40] H. L. F. v. Helmholtz, *On the Sensations of Tone as a Physiological Basis for the Theory of Music*, A J. Ellis (Trans.) (Longmans, Green, and Co., London, UK, 1895). https://archive.org/details/onsensationsofto00helmrich.

[41] M. Wertheimer, "Laws of Organization in Perceptual Forms," in W. D. Ellis (Ed.), *A Source Book of Gestalt Psychology*, pp. 71–88 (Kegan Paul, Trench, Trubner & Company , London, UK, 1938). https://doi.org/10.1037/11496-005.

[42] R. Rasch, R. Plomp, "The perception of musical tones," in D. Deutsch (Ed.), *The Psychology of Music*, pp. 89–112 (Academic Press, Cambridge, MA, 1982), https://doi.org/10.1016/B978-012213564-4/50005-6.

[43] B. C. Moore, B. R. Glasberg, R. W. Peters, "Thresholds for Hearing Mistuned Partials as Separate Tones in Harmonic Complexes," *J. Acoust. Soc. Am.*, vol. 80, no. 2, pp. 479–483 (1986 Aug.). https://doi.org/10.1121/1.394043.

[44] Vitalic, "Poney Part 1," on *OK Cowboy*, Different Records/PIAS (2005 Apr.).

[45] J. L. Goldstein, "An Optimum Processor Theory for the Central Formation of the Pitch of Complex Tones," *J. Acoust. Soc. Am.*, vol. 54, no. 6, pp. 1496–1516 (1973 Dec.). https://doi.org/10.1121/1.1914448.

[46] J. C. R. Licklider, "A Duplex Theory of Pitch Perception," *J. Acoust. Soc. Am.*, vol. 23, no. 1 Supplement), paper. 147 (1951 Dec.). https://doi.org/10.1121/1.1917296.

[47] L. Rabiner, M. Cheng, A. Rosenberg, C. McGonegal, "A Comparative Performance Study of Several Pitch Detection Algorithms," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 24, no. 5, pp. 399–418 (1976 Oct.). https://doi.org/10.1109/TASSP.1976.1162846.

[48] J. F. Schouten, "The Residue, a New Component in Subjective Sound Analysis," in *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen, 1940*, vol. 43, pp. 356–365 (Amsterdam, the Netherlands) (1940 Feb.). https://dwc.knaw.nl/DL/publications/PU00017418.pdf.

[49] E. De Boer, "On the "Residue" and Auditory Pitch Perception"," in *Auditory System: Clinical and Special Topics*, *Handbook of Sensory Psychology*, vol. 5/3, pp. 479–583 (Springer, Berlin, Germany, 1976). https://doi.org/10.1007/978-3-642-66082-5_13.

[50] H. Fletcher, W. A. Munson, "Loudness, Its Definition, Measurement and Calculation," *Bell Syst. Tech. J.*, vol. 12, no. 4, pp. 377–430 (1933 Oct.). https://doi.org/10.1002/j.1538-7305.1933.tb00403.x.

[51] D. W. Robinson, R. S. Dadson, "A Re-Determination of the Equal-Loudness Relations for Pure Tones," *Br. J. Appl. Phys.*, vol. 7, no. 5, p. 166 (1956 May). https://doi.org/10.1088/0508-3443/7/5/302.

[52] E. Skovenborg, S. H. Nielsen, "Evaluation of Different Loudness Models With Music and Speech Material," presented at the *117th Convention of the Audio Engineering Society* (2004 Oct.), paper 6234. https://aes2.org/publications/elibrary-page/?id=12891.

[53] ISO, "Acoustics—Normal Equal-Loudness Level Contours," Standard 226:2023 (2023 Mar.). https://www.iso.org/standard/83117.html.

[54] F.-R. Stöter, S. Uhlich, A. Liutkus, Y. Mitsufuji, "Open-Unmix – A Reference Implementation for Music Source Separation," *J. Open Source Softw.*, vol. 4, no. 41, p. 1667 (2019 Sep.). https://doi.org/10.21105/joss.01667.

[55] G. Burns, "A Typology of 'Hooks' in Popular Records," *Popular Music*, vol. 6, no. 1, pp. 1–20 (1987 Jan.). https://www.tagg.org/xpdfs/burns87.pdf.

[56] F. Bimbot, E. Deruty, G. Sargent, E. Vincent, "System & Contrast: A Polymorphous Model of the Inner Organization of Structural Segments Within Music Pieces," *Music Percept.*, vol. 33, no. 5, pp. 631–661 (2016 Jun.). https://doi.org/10.1525/mp.2016.33.5.631.

[57] D. R. Grimes, "String Theory-The Physics of String-Bending and Other Electric Guitar Techniques," *PloS One*, vol. 9, no. 7, paper e102088 (2014 Jul.). https://doi.org/10.1371/journal.pone.0102088.

[58] Pink Floyd, "Shine On You Crazy Diamond - Live at Knebworth," (1990 Jun.), DavidGilmourHD. https://youtu.be/cRx5LYWQxoI&t=62.

[59] Vitalic, "Vitalic Live at Rotonde Stalingrad in Paris, France for Cercle," Cercle (2017 Nov.). https://youtu.be/IGUAChyue2I.

[60] E. Deruty, M. Grachten, "'Melatonin': A Case Study on AI-induced Musical Style," in *Proceedings of the 3rd Conference on AI Music Creativity* (online) (2022 Sep.). https://doi.org/10.5281/zenodo.7088302.

[61] M. Grachten, E. Deruty, A. Tanguy, "Auto-Adaptive Resonance Equalization Using Dilated Residual Networks," in *Proceedings of the 20th International Society for Music Information Retrieval Conference*, pp. 405–411 (Delft, Netherlands) (2019 Nov.). https://doi.org/10.5281/zenodo.3527830.

[62] M. Mersenne, *Harmonie Universelle* (Cramoisy et Ballard, Paris, 1636).

[63] R. S. Turner, "The Ohm-Seebeck Dispute, Hermann von Helmholtz, and the Origins of Physiological Acoustics," *Br. J. Hist. Sci.*, vol. 10, no. 1, pp. 1–24 (1977 Mar.). https://www.jstor.org/stable/pdf/4025578.pdf.

[64] B. C. Moore, *An Introduction to the Psychology of Hearing* (Emerald Group Publishing Limited, Bingley, UK, 2012).

[65] N. H. Fletcher, "Nonlinear Dynamics and Chaos in Musical Instruments," in D. Green and T Bossomaier (Ed.), *Complex Systems: From Biology to Computation*, pp. 106–117 (IOS Press, Amsterdam, Netherlands, 1993).

[66] W. Dixon Ward, "Musical Perception," in J. Tobias (Ed.), *Foundations of Modern Auditory Theory*, vol. 1, pp. 405–446 (Academic Press, New York, NY, 1970).

[67] R. Plomp, "The Ear as a Frequency Analyzer," *J. Acoust. Soc. Am.*, vol. 36, no. 9, pp. 1628–1636 (1964 Sep.). https://doi.org/10.1121/1.1919256.

[68] J. P. Rameau, "Chap. XIX. Suite du Chapitre précédent, où il paroît que la Melodie provient de l'Harmonie," in *Traité de l'Harmonie Réduite à ses Principes Naturels*, pp. 138–140 (Jean-Baptiste Ballard, Paris, France, 1722).

[69] C. Dahlhaus, in *Studies on the Origin of Harmonic Tonality*, p. 72 (Princeton University Press, Princeton, NJ, 1990).

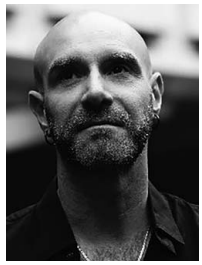
[70] A. Cœurdevey, in *Histoire du Langage Musical Occidental* , p. 82 (Presses Universitaires de France, Paris, France, 1998).

[71] C. E. Cancino-Chacón, M. Grachten, W. Goebl, G. Widmer, "Computational Models of Expressive Music Performance: A Comprehensive and Critical Review," *Front. Digit. Humanit*., vol. 5, paper 25 (2018 Oct.). https://doi.org/10.3389/fdigh.2018.00025.

THE AUTHORS



Emmanuel Deruty



Pascal Arbez-Nicolas



David Meredith

Emmanuel Deruty is a researcher specializing in music production and its relations to technology and science. He is part of the Music Project at Sony Computer Science Laboratories (CSL), which focus on artificial intelligence (AI)–assisted music production.

•

Pascal Arbez-Nicolas, known professionally as Vitalic, is a French electronic music producer who blends techno, electro, and house with classical, rock, and 1970s electronic music. His career reflects a continuous evolution of electronic music styles while maintaining a rebellious and experimental edge.

•

David Meredith is an associate professor in the Department of Architecture, Design, and Media Technology at Aalborg University in Denmark. His research focuses on music information retrieval, mathematical music theory, computational musicology, and pattern matching.