OPEN ACCESS Freely available online

E. Deruty, D. Meredith, M. Grachten, P. Arbez-Nicolas,
A. H. Jørgensen, O. S. Hansen, C. N. Petersen, and M. Stensli,
"Methods for Pitch Analysis in Contemporary Popular
Music: Multiple Pitches From Harmonic Tones in Vitalic's Music,"
J. Audio Eng. Soc., vol. 74, no. 5, pp. 270–287 (2026 May).
https://doi.org/10.17743/jaes.2022.0264.

Methods for Pitch Analysis in Contemporary Popular Music: Multiple Pitches From Harmonic Tones in Vitalic's Music

EMMANUEL DERUTY,[1, 2, *] **DAVID MEREDITH,**[2]
(emmanuel.deruty@sony.com) (dave@create.aau.dk)

MAARTEN GRACHTEN,[3] **PASCAL ARBEZ-NICOLAS,**[4]
(maarten@grachten.eu)

ANDREAS HASSELHOLT JØRGENSEN,[2] **OLIVER SØNDERMØLLE HANSEN,**[2]
(ajorge22@student.aau.dk) (osha21@student.aau.dk)

CHRISTIAN NØRKÆR PETERSEN,[2] **AND MAGNUS STENSLI**[2]
(cpeter22@student.aau.dk) (mstens17@student.aau.dk)

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

This study shows that contemporary popular music with electronic elements can intentionally exploit multiple or ambiguous pitch perceptions arising from a single harmonic complex tone. The phenomenon is illustrated with examples from the work of electronic artist Vitalic and from widely used electronically mediated tones. Two listening tests were conducted: (1) evaluation of the number of simultaneous pitches perceived in sequences of quasi-harmonic tones, and (2) manual pitch transcription of selected sequences. Relationships between signal characteristics and pitch perception were subsequently analyzed. In the musical sequences under consideration, the synthesized harmonic tones were found to convey a greater number of perceived pitches than the acoustic tones. Multiple ambiguous pitches were associated with features such as prominent upper partials and specific autocorrelation profiles. In contemporary popular music with an electronic component, harmonic tones can convey multiple ambiguous pitches. The set of perceived pitches depends on both the listener and the listening conditions.

Keywords: electronic music, listening tests, music production, pitch perception

1 INTRODUCTION

The human ability to perceive distinct pitches simultaneously from a single harmonic complex tone—referred to here as *tonal fission*—has long been debated [1] and demonstrated using artificial stimuli [2–4]. However, tonal fission has generally not been considered significant in everyday musical listening, due to the specific conditions thought necessary for it to occur [5, 6, p. 438]. Harmonic complex tones produced by musical instruments are typically considered to transmit a single pitch, corresponding to the fundamental frequency (f_0)—the lowest harmonic of the tone.

Nevertheless, the phenomenon plays a documented role in music. The deliberate elicitation of multiple pitches from a single quasi-harmonic tone appears in non-Western traditions such as overtone singing [7], as well as in contemporary classical and jazz music, where it is referred to as *multiphonics* [8, 9] and employed as a compositional or expressive technique.

In contemporary popular music (CPM), as defined by Deruty et al. [10], the phenomenon of tonal fission has remained largely undocumented. Only recently has it been identified as a stylistic element in the works of contemporary electronic music producers such as Hyper Music [11–13] and Vitalic [14]. These studies make two key observations. First, the ubiquitous use of synthesizers and effects processors gives artists a high degree of control over

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

sound, allowing them to create quasi-harmonic tones that are ambiguous in pitch content. Second, Deruty et al. [13, 14] find that the artists they studied use tonal fission and pitch ambiguity deliberately as compositional tools—traits that may be considered idiomatic to the technology in the sense defined by Huron and Berec [15].

Such practices challenge the notion that each harmonic complex tone is associated with a single, intrinsic pitch—a view widely held within the music information retrieval community, particularly in the context of pitch-tracking tasks [16]. They also prompt a reevaluation of ambiguity as an integral component of musical discourse, a trait already noted in earlier literature [17].

Whereas prior work by Deruty et al. [13, 14] draws on producers' perspectives and signal analysis to question the link between a harmonic tone and a single pitch, the present study approaches this issue through listening tests focused on tonal fission. The listening tests use excerpts from the work of electronic music producer Vitalic (with whom the authors were in contact during the development of the study), along with widely employed sound sources representative of other genres, such as the distorted electric guitar and TR-808-style bass.

The choice of Vitalic as the main case study reflects both methodological and practical considerations. Methodologically, the study aims to demonstrate and characterize tonal fission in detail rather than to survey its prevalence; a single, thoroughly examined body of work better serves this objective. Practically, Vitalic's direct participation as a coauthor provided access to compositional intent, production details, and private interviews—a level of collaboration that is rare with commercially active artists and cannot be assumed to be readily available with other producers. His explicit awareness of his own use of tonal fission [18] further motivated this choice. Although prior work [11–13] suggests that the phenomenon extends to other artists, the decision was made to prioritize analytical depth over breadth.

The results suggest that single quasi-harmonic complex tones in these contexts can give rise to multiple pitch sensations, with perceived pitches varying across listeners. The paper is structured as follows. SEC. 2 presents the context of the study, and SEC. 3 outlines the methods. SECS. 4 and 5 detail the two listening tests and their respective results. SEC. 6 explores links between a tone's signal and the phenomenon of tonal fission. SEC. 7 compiles listener comments to provide further insight into the test results. Broader perspectives are discussed in SEC. 8, and conclusions are presented in SEC. 9[1].

2 BACKGROUND

This section introduces the notion of *tonal fission*, in contrast to *tonal fusion*. It outlines how Vitalic's music production deliberately employs tonal fission for expressive purposes and defines several useful concepts related to pitch perception.

2.1 Tonal Fusion and Fission

The term *tonal fusion* refers to the tendency for certain combinations of tones to be perceived as a single auditory image [19, 20]. Typically, tones separated by intervals with strong harmonic relationships are more likely to produce tonal fusion across listeners. In a distinct context, the term *fission* has been used to describe the perception of a single tone sequence as originating from two or more sources [21–23].

The samples examined in this study exhibit fission arising from a single tone, a phenomenon referred to here as *tonal fission*. It relates to the well-documented ability of humans to perceive multiple distinct pitches simultaneously from a single (quasi-)harmonic complex tone—a topic that has long been the subject of research and debate [1, 24–28]. Dewitt and Crowder [20] also suggest a conceptual link between tonal fusion and tonal fission, hypothesizing that the pitch ambiguity observed in the fusion of two harmonic tones could likewise arise from a single harmonic tone.

2.2 Vitalic's Perspective: Tonal Fission as a Musical Parameter

A notable aspect of the work of electronic music producer, Vitalic, is what he describes as "a melody inside the bass" [18], referring to a secondary melodic line embedded within a sequence of distinct quasi-harmonic tones. More broadly, Vitalic has confirmed in private interviews that he treats the number of perceived pitches arising from a single tone as a musical parameter and that these pitches are often intentionally ambiguous. This approach is not unique to his practice, as suggested by studies involving other music producers and examples from the broader music market [11–14, 29].

2.3 Terminology

Regarding tones, the following terminology is used:

1. *Partial*: the spectral representation of a single periodic sine wave, corresponding to what von Helmholtz [26] describes as a *constituent* or *partial tone*.
2. *Harmonic complex tone*: an ensemble of partials that are integer multiples of a fundamental frequency (f_0), as defined by von Helmholtz [26] as a *compound*. When the partial corresponding to f_0 is absent ("missing fundamental"), it may still be determined as the greatest common divisor of the partials' frequencies.
3. *Inharmonic tone*: a tone that is not harmonic. Tones can vary in their degree of inharmonicity, allowing for the identification of quasi-harmonic tones, in which each partial of the least deviating harmonic series can be clearly associated with a partial of the original tone. In this case, f_0 may be evaluated as the

[1] Supplementary material with audio examples is available at http://vhtmp-suppl-mat.s3-website.eu-west-3.amazonaws.com/.

fundamental frequency of the least deviating harmonic series [30].

4. *Overtone* or *upper partial*: refers to any partial in a tone other than f_0. This aligns with von Helmholtz's definition of *upper partial tones* [26].
5. *Harmonics*: defined by von Helmholtz [26] as the sinusoidal components (partials) of a harmonic complex tone, where the nth harmonic has a frequency n times the fundamental. In the case of quasi-harmonic tones, the nth harmonic's frequency neighbors the nth multiple of the f_0 of the least deviating harmonic series.

2.4 Pitch Perception: Spectral and Temporal Modeling

Yost [4] distinguishes between *spectral* and *temporal* models of pitch perception, whose main proponents are Goldstein [31] and Licklider [32], respectively. In spectral modeling, pitch is extracted from spectral components that are resolved by the auditory periphery—that is, components whose frequency spacing exceeds the width of critical bands. In contrast, temporal modeling typically derives pitch from the autocorrelation of the signal. Depending on the tone, the spectral and temporal approaches may either converge or diverge.

In this paper, simplified forms of both approaches are applied to link perceived pitches with measurable signal characteristics. A simplified version of spectral modeling is used to associate perceived pitches with individually audible upper partials. While temporal modeling often relies on the highest autocorrelation peak beyond zero lag [33, p. 401], frequency differences between adjacent partials are also considered as indicators of temporal structure.

3 METHODS

This section describes the methods used in the study, including audio source separation, equal-loudness-contour weighting, and comparative interpretation of signal-based and listening-based analyses. It then outlines the prior methodology that led to the present investigation.

3.1 Source Separation

The signal analyses and listening tests in this paper focus on isolated tone sequences. When such tones cannot be directly obtained from the original tracks, they are extracted through source separation using the X-UMX algorithm [34]. This algorithm separates audio into six categories: drums, guitar, bass, piano, vocals, and other. The sum of the separated files exactly reproduces the original.

Vitalic's bass tracks often differ from traditional electric bass, leading the algorithm to classify the lower bass spectrum as bass and the upper spectrum as other. Similarly, the lower range of harmonic synthesizers may also be categorized as bass. This ambiguity is reflected in some samples used in the study.

3.2 Psychoacoustic Weighting

Signal analyses in this study incorporate equal-loudness-contour weighting, recognizing that humans are not equally sensitive to all frequencies [35]. Various models exist that weight the power spectrum to reflect perception [35–37]. One such model is ISO 226:2023 [38]. The ISO 226:2023 50-phon equal-loudness contour is applied before spectral analysis to produce signal analysis results that better reflect what listeners actually hear.

Applying such a model prior to signal analysis is essential, as some f_0 values are very low. In SEC. 5, two frequencies are near 50 Hz, and one is near 32 Hz, corresponding to attenuations of about 40 and 50 dB relative to 1,000-Hz tones. These levels exceed the typical interlistener differences that correspond to severe hearing impairment [39].

3.3 Signal-Based and Listening-Based Comparative Interpretation

SEC. 6 relates the pitches perceived during the listening tests to features derived from the samples' short-time Fourier transform (STFT) and short-time autocorrelation (STAC). This approach, previously applied in [13] and [14], serves two main objectives: (1) to account for why particular pitches were perceived and (2) to propose an alternative to the score for addressing the volatility of sounds, an issue recognized since the 14th century [40] and further examined in [13].

3.4 Prior 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 equal temperament. Observations were sorted into six nonexclusive categories:

1. Presence of complex instrumental tones.
2. Presence of complex vocal tones (e.g., vocoder).
3. Tuning not consistent with equal temperament.
4. Presence of continuous pitch trajectories.
5. Presence of low pitch-strength instruments (e.g., drums).
6. Presence of flanger/chorus resulting in additional pitch.

The present study focuses on the first category. Of the 72 songs examined, 38 feature instrumental tracks based on tones that evoked more than one pitch to the authors. In most cases, these tracks are prominent and use such tones throughout the entire song. Therefore, while the study's conclusions do not apply to all of Vitalic's work, they are representative of over half his discography.

Table 1 lists the extracts used in this study. Nine extracts from Vitalic's music were selected for Listening Test 1 (SEC. 4), with three also used in Listening Test 2

Table 1. List of audio samples used in the study.

Sample	Author	Title	Reference	Timing	Instrument	Structure	Separation
1	Vitalic	"No Fun"	[69]	0:12–0:14	Main synthesizer	Inharmonic	Yes
2	Vitalic	"Forgiven"	[70]	0:00–0:01	Bass	Quasi-harmonic	Yes
3	Vitalic	"La Mort sur le Dance Floor"	[71]	0:03–0:04	Bass	Quasi-harmonic	No (solo)
4	Vitalic	"No Fun," top line, transcribed for piano	[69]	0:12–0:14	Piano	Quasi-harmonic	No (solo)
5	Vitalic	"Sweet Cigarette"	[72]	0:29–0:31	Main synthesizer/bass	Quasi-harmonic	Yes
6	B. Tchaikovsky	"Solo Cello Suite in D Minor: III. Aria. Adagio"	[73]	0:15–1:38	Cello	Quasi-harmonic	No (solo)
7	Vitalic	"Stamina"	[74]	0:40–0:47	Main synthesizer/bass	Quasi-harmonic	Yes
8	Vitalic	"Next I'm Ready"	[75]	0:52–0:59	Bass	Quasi-harmonic	Yes
9	Vitalic	"La Mort sur le Dance Floor"	[71]	0:04–0:08	Bass	Quasi-harmonic	No (solo)
10	J.S. Bach	"Partita II BWV 1004" in D minor: V. Ciaccona"	[76]	1:34–1:58	Violin	Quasi-harmonic	No (solo)
11	Vitalic	"See the Sea (Blue)"	[77]	0:42–0:46	Main synthesizer/bass	Quasi-harmonic	Yes
12	Vitalic	"No Fun"	[69]	0:30–0:38	Main synthesizer	Inharmonic	Yes
13	...	...	...	...	Guitar power chord	Quasi-harmonic	No
14	...	...	...	...	808 Woofer Warfare, mode 1	Quasi-harmonic	No

(SEC. 5). The primary selection criterion was ensuring diversity within a small sample pool. Listening Test 1 also includes three monodic sequences from acoustic instruments, serving as a control group. One of these sequences was included in Listening Test 2, which also features a guitar power chord and a TR808 bass-style sample. These two samples were included to suggest that the study's observations may extend to different music genres.

4 LISTENING TEST 1: HOW MANY SIMULTANEOUS PITCHES?

4.1 Procedure

Listening Test 1 included samples 1–12. For each extract, the 59 participants reported the number of simultaneous pitches they perceived. The questionnaire collected data on age (70% of respondents were under 35 years old), hearing condition, musical and audio expertise, and listening environment and monitoring system. After the test, five nonexpert listeners were interviewed about their responses (see SEC. 7).

Listening conditions were deliberately not controlled, in order to reflect the participants' typical listening environments. Enforcing a single set of conditions would have biased the experiment toward those specific settings. This decision is further supported a posteriori, as SEC. 5 suggests that the perceived pitches may vary depending on the listening conditions.

A preliminary experiment indicated a habituation phenomenon: the more listeners focused on the number of pitches, the more they perceived. One participant took the test twice, reporting a mean of 2.4 simultaneous pitches the first time and three the second time. This observation is coherent with Nunes et al.'s remark [41], according to which lexical repetition facilitates ease of processing. To mitigate this effect, the test was offered in four versions with different extract orders. One version is accessible at https://golisten.ucd.ie/task/acr-test/6746160c75e8b306f88e34bd.

4.2 Results

Fig. 1 summarizes the experimental results, with extracts ordered by increasing number of perceived simultaneous pitches. Not a single sample—acoustic or electronic—elicits a unanimous perception of a single pitch. The selected Vitalic extracts mostly lead to the perception of multiple pitches, with the mean number of perceived pitches per excerpt ranging from 1.6 to 2.5. The two samples with inharmonic tones (see SEC. 6.4) rank higher than most, though some quasi-harmonic tones yield comparable results, and one even exceeds them.

4.2.1 Sound Categories

Based on the initial expectation that acoustic instrument sounds elicit single-pitch perception, the authors distinguish between acoustic sounds and electronic sounds. Given that inharmonic partials may be perceived as distinct pitches [22, 42], electronic sounds are further subdivided into quasi-harmonic and inharmonic categories.

Fig. 2 shows the average number of perceived pitches as a function of harmonicity and musical expertise. Both electronic categories are associated with significantly higher odds of tonal fission compared to the acoustic samples, with the strongest effect observed for the two inharmonic electronic sounds.

4.2.2 Source Separation

Table 1 shows that several samples were obtained through source separation. This is unlikely to affect the results, as the synth bass parts are highly prominent and the remaining elements of the mix consist primarily of drums, which are relatively easy to separate.

4.2.3 Listener Characteristics

A preliminary Pearson correlation analysis reveals moderate associations among age, musical knowledge, and audio expertise. Age group shows a moderate positive correlation with musical knowledge ($r = 0.44$) but only a weak

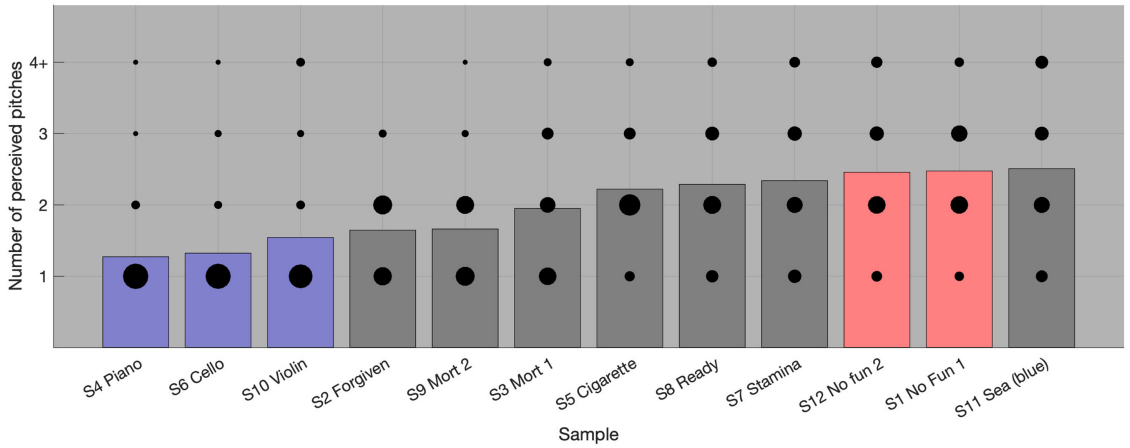


Fig. 1. Number of perceived pitches. See Supplementary Material[1] SEC. A for the corresponding audio. Dot size areas indicate the number of responses. The bar graphs represent the overall mean of the responses. The blue bars correspond to acoustic sources, and the red bars correspond to sequences of inharmonic tones.

correlation with audio knowledge ($r = 0.19$). Meanwhile, musical knowledge and audio knowledge demonstrate a strong positive correlation ($r = 0.66$), suggesting substantial overlap between these two areas of expertise.

The effect of musical and audio expertise on the number of perceived pitches is examined. As illustrated in Fig. 2 and Supplementary Material[1] SEC. A, perceiving only a single pitch per tone occurs exclusively among audio and music experts listening to acoustic instruments.

A logistic regression analysis, treating the number of perceived pitches as an ordinal variable and controlling for age, listening device, and environment, reveals a significant negative relationship between musical expertise and pitch multiplicity ($p < 0.001$), indicating that greater musical training is associated with fewer perceived pitches. In contrast, audio expertise shows no significant effect ($p = 0.209$).

4.2.4 Summary

Many subjects do perceive multiple simultaneous pitches in the presented harmonic complex tones, especially for the synthetic examples. SEC. 7 elaborates on the subjects' perceptions of the tones.

5 LISTENING TEST 2: TRANSCRIPTIONS

Listening Test 1 showed that quasi-harmonic complex tones may be widely perceived as transmitting multiple perceived pitches. For Listening Test 2, tonal fission was focused on in more detail. Participants with high musical expertise from the first test were selected. These participants were asked to transcribe the pitches they heard.

5.1 Procedure

Seven participants transcribed the pitches they perceived in a smaller set of six samples. All participants used professional listening systems in a quiet room. The reduced number of samples compared to Listening Test 1 reflects the time-intensive nature of the task. Four samples were selected from Listening Test 1 for their diversity: one acoustic tone (sample 6), two quasi-harmonic electronic tones (samples 2 and 7), and one inharmonic tone (sample 1) for comparison. Samples 13 (power chord) and 14 (808-type bass drum) were added as representatives of specific musical genres—rock and urban music.

In this experiment, listeners used a piano to match the pitches in the samples. They were also asked to indicate

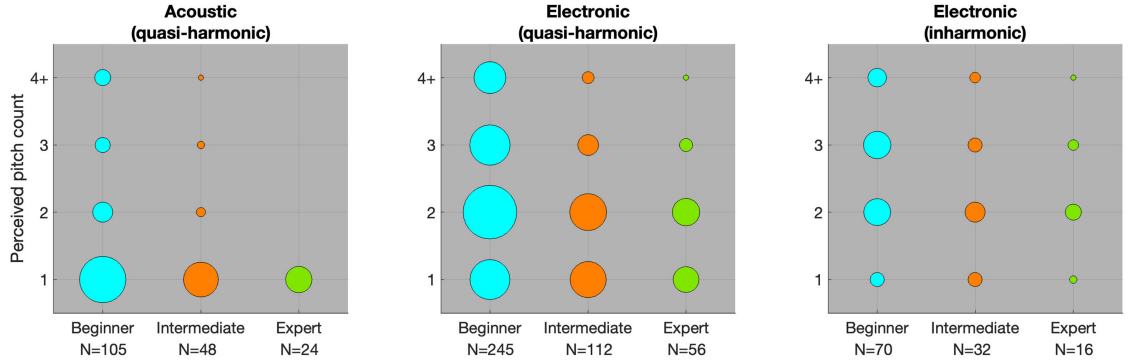


Fig. 2. Distribution of perceived pitch counts by musical expertise and sound category. Dot areas indicate the number of occurrences. The dataset includes 59 listeners and 12 sound samples, totaling 708 individual listenings.

whether a tone sounded out of tune and, if so, to estimate the degree of detuning. Although the standard pitch-matching procedure involves presenting participants with a stimulus alongside a sine wave of adjustable frequency—with responses recorded as the adjusted frequency once the perceived pitches are matched [43]—the authors opted for this more familiar setup despite its lower precision, as it is more intuitive for individuals with musical training.

Listeners rated the certainty of each perceived pitch. A certainty of 1 indicates pitches that are consistently and effortlessly perceived, while low values (e.g., 0.2) reflect pitches that require focused attention or are subject to variation across listening conditions. In samples 13 and 14, where no discrete pitch change occurred, listeners were instructed to align their responses to a 0.5-s temporal grid during transcription. The results were discussed both qualitatively (SEC. 5.2) and quantitatively (SEC. 5.3) and related to signal characteristics (SEC. 6).

5.2 Results

5.2.1 Cello Sequence (Sample 6)

See Fig. 3 for a compilation of the transcriptions. In line with the results of Listening Test 1, listeners were unanimous in their transcription of the cello example. For each note in the sequence, all listeners identified the same single pitch with maximum certainty.

5.2.2 "Forgiven" (Sample 2) and "Stamina" (Sample 7)

Unlike the cello transcriptions, (1) listeners generally perceived multiple pitches from the tone, and (2) different listeners may perceive different pitches. These pitches are either melodically close to the f_0 (around G♯1 in both cases) or harmonically related to it in the sense of Ptolemy [44].

5.2.3 "No Fun" (Sample 1)

Listener consensus on a single dominant pitch is higher than for samples 2 and 7. This contrasts with traditional views of pitch perception, where quasi-harmonic tones (samples 2 and 7) are typically associated with a single perceived pitch, while inharmonic tones (this sample) are expected to prompt multiple perceived pitches. In this sample, listeners unanimously identified a main melodic line (B3-F♯4-B3-F♯4-G4-A4-A4-F♯4). Some also perceived a parallel melodic line one octave below, while one listener reported hearing a similar line one octave above. Several listeners also perceived a constant low pitch (B2).

5.2.4 Power Chord (Sample 13)

Power chords are a defining feature of many rock styles. Although they are typically produced by simultaneously playing two or three notes on a guitar (root, fifth, and optionally the octave), the result following the distortion stage is a single harmonic complex tone—see SEC. 6.5. The findings exhibit several parallels with those of samples 2 and 7: multiple pitches perceived per tone, pitch perception varying between listeners, and perceived pitches that are melodically and/or harmonically related to the f_0 (E1).

5.2.5 808 Woofer Warfare, Mode 1 (Sample 14)

Bass-drum tones produced by the Roland-808, a key element in many electronic and urban styles, are harmonic complex tones—see SEC. 6.6 and [45]. The results resemble those of samples 2, 7, and 13: multiple pitches per tone, listener-dependent pitch perception, perceived pitches either near or harmonically related to the expected pitch, and octave ambiguity. The range of responses is broad, spanning over five octaves. One listener perceived a C1, an octave below the f_0 (C2).

5.3 Quantitative Evaluation

This evaluation aims to determine to which categories of partials the perceived pitches correspond. It also seeks to quantify the degree of listener consensus.

To assess partial categories, the proportion of perceived pitches corresponding to (1) the f_0, (2) harmonics—particularly octaves, and (3) nonharmonic frequencies is measured. In this framework, "No Fun" (sample 1) is treated as a special case due to its inharmonic tones. For this sample, the f_0 is defined as the frequency spacing between consecutive partials (see SEC. 6.4).

To quantify consensus, using variance is avoided, as frequency differences are not inherently metric. For example, in harmonic distance (in the sense of Winter et al. [46]), 200 Hz may be closer to 100 Hz than to 160 Hz. Instead, the authors use the weighted Jaccard similarity (WJS), which measures the degree to which the pitch sets—along with their associated certainty values—transcribed by different listeners coincide for a given tone. Let A and B be the sets of pitches transcribed by two listeners, and let $c_A(p)$ and $c_B(p)$ return the certainty assigned to pitch p by each listener, with $c_X(p) = 0$ when $p \notin X$. The WJS is then defined as

$$WJS(A, B) = \frac{\sum_{p \in A \cap B} \min(c_A(p), c_B(p))}{\sum_{p \in A \cup B} \max(c_A(p), c_B(p))}. \quad (1)$$

Table 2 presents the evaluation results, while Fig. 4 provides a visual representation. Several observations can be made:

- For sample 6 (cello), all values indicate unanimous pitch perception across listeners, suggesting a clear and stable pitch is conveyed.
- Samples 2 and 7 (synthesizer, quasi-harmonic) produce similar outcomes: a lack of consensus among listeners, with most perceived pitches corresponding to octaves of the f_0. This suggests that pitch ambiguity is concentrated around octave-related partials.
- Sample 7, which involves inharmonic tones, surprisingly yields higher consensus than the nonacoustic harmonic tones. The pitch corresponding to the difference between partials—potentially perceivable through temporal modeling—is never reported. However, upper octaves account for a substantial proportion of the perceived pitches. The prevalence of harmonics in listener responses may be partly explained by the high density of partials in the register where these pitches occur.

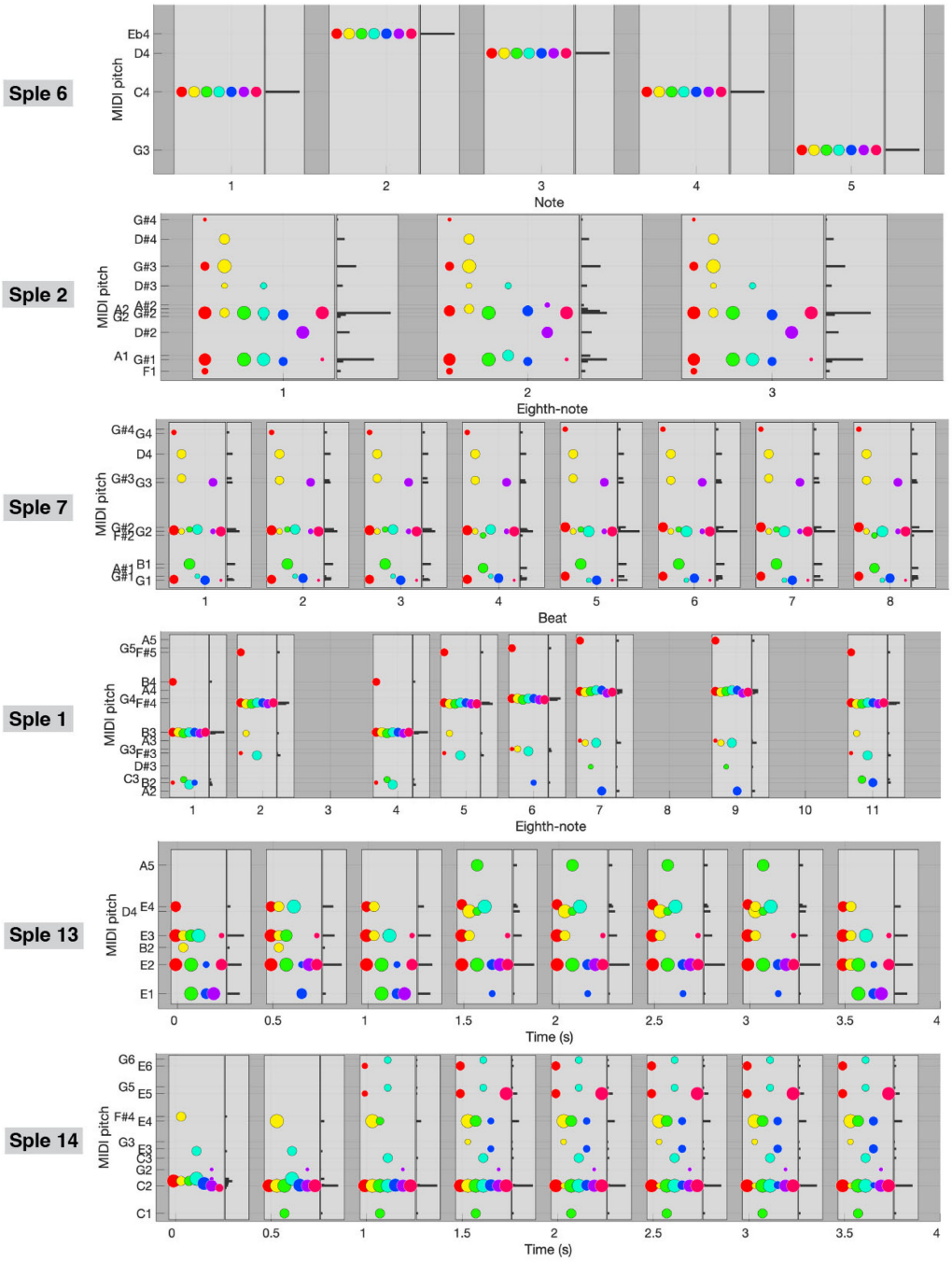


Fig. 3. Transcribed pitches for the six samples. Each participant is associated with one color and one vertical position inside the clear rectangles. The area of each dot denotes certainty. The right-hand bar graphs show the distribution across subjects.

Table 2. Value outputs for the evaluation criteria. Proportion of perceived pitches corresponding to the f_0.

Sample	Fundamental	Harmonics All (Octaves)	Others	WJS Mean (Standard Deviation)
Cello (sample 6)	1.00	0.00 (0.00)	0.00	1.00 (0.00)
“Forgiven” (sample 2)	0.31	0.61 (0.52)	0.08	0.20 (0.08)
“Stamina” (sample 7)	0.21	0.68 (0.60)	0.11	0.09 (0.01)
“No Fun” (sample 1)	0.00	0.93 (0.41)	0.07	0.62 (0.10)
Power chord (sample 13)	0.12	0.88 (0.76)	0.00	0.22 (0.04)
808 Woofer Warfare (sample 14)	0.55	0.36 (0.03)	0.08	0.38 (0.16)

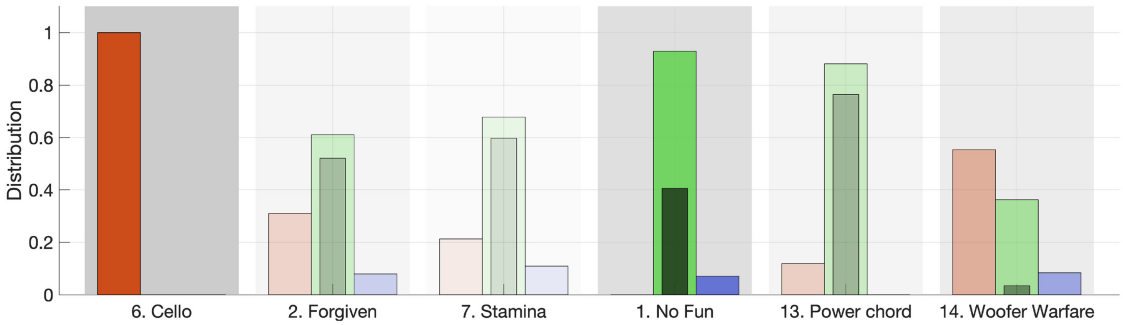


Fig. 4. Summary of quantitative evaluation. For each sample, the leftmost bar (orange) shows the proportion of perceived pitches corresponding to the f_0; the middle bar (green) corresponds to harmonics, with the darker segment indicating the octaves; and the rightmost bar corresponds to other partials. The transparency of each group reflects the WJS.

- In sample 13 (power chord), the majority of perceived pitches derive from upper octaves.
- Sample 14 stands out among nonacoustic quasi-harmonic tones. It exhibits strong perception of the fundamental frequency and very low octave ambiguity. Rather than the octave, other harmonics appear to be more salient in listeners' perceptions.

6 LISTENING TEST 2 RESULTS AND SAMPLE SIGNALS

Harmonic complex tones are commonly associated with a single pitch, yet tonal fission was observed to a notable extent in Listening Test 2. This contrast calls for a closer look at the samples to explore potential links between this phenomenon and signal features.

6.1 Cello Sequence (Sample 6)

Fig. 5 shows the STFT and STAC of sample 6, illustrating a case in which tonal fission is not observed. The f_0 and the first peak of the autocorrelation function after the zero crossing align and correspond to the perceived pitches. Both spectral and temporal modeling lead to the same pitch. These observations are consistent with established findings on pitch perception in harmonic complex tones (e.g., [31, 32, 47, 48]).

6.2 "Forgiven" (Sample 2)

6.2.1 Sample Details

Sample 2 was produced using XferRecords' Serum, a software wavetable synthesizer. As shown in Fig. 6(a), the sample consists of three consecutive quasi-harmonic tones with an f_0 near G♯1. All partial frequencies decrease slightly and gradually throughout each tone.

6.2.2 Perceived Pitches and Signal

Fig. 6(a) shows that most perceived pitches align with partials, suggesting spectral modeling. G♯1 (f_0) and G♯2 (harmonic 1) are also consistent with temporal modeling. Indeed, the first and third tones primarily contain odd harmonics and exhibit a relatively weak fundamental, resembling harmonic complexes with missing lower partials, as described by Yost [4]. In such cases, pitch perception becomes ambiguous—corresponding either to the f_0 [see Fig. 6(b), autocorrelation peak at G♯1] or the spacing between partials. This spacing is reflected in the first minimum of the autocorrelation function following the initial peak [14], which in Fig. 6(b) corresponds to G♯2. One perceived pitch remains unexplained: D♯2, approximately one octave below the fifth harmonic.

6.3 "Stamina" (Sample 7)

6.3.1 Sample Details

Sample 7 was created using a Novation Bass Station, a 1993 analogue synthesizer. It features a quasi-harmonic tone with an f_0 near G1 and attenuated partials above the fifth harmonic. Partial frequencies are modulated conjointly, while amplitudes are modulated individually. Above the fourth partial, odd and even harmonics alternate. As shown in Fig. 7(b), modulation of partial frequencies produces several closely spaced autocorrelation peaks that evolve periodically.

6.3.2 Perceived Pitches and Signal

Fig. 7(a) illustrates the relatively low loudness of the lower harmonics, suggesting that the perception of low pitches (around G1 and G2) relies on temporal modeling. In contrast, the alignment between higher perceived pitches and upper partials indicates a reliance on spectral modeling. The continuous modulation of partial frequencies may account for the variety of perceived tunings.

6.4 "No Fun" (Sample 1)

6.4.1 Sample Details

The main synthesizer part in "No Fun" was built from tones sampled from the Roland VariOS unit. It consists of a sequence of inharmonic tones exhibiting what Yost [4] describes as "pitch-shift of the residue." Deruty [29] provides a detailed analysis of the sample. The frequency differences between adjacent partials remain approximately constant throughout the extract (around B1), even though the partials themselves are shifted.

6.4.2 Perceived Pitches and Signal

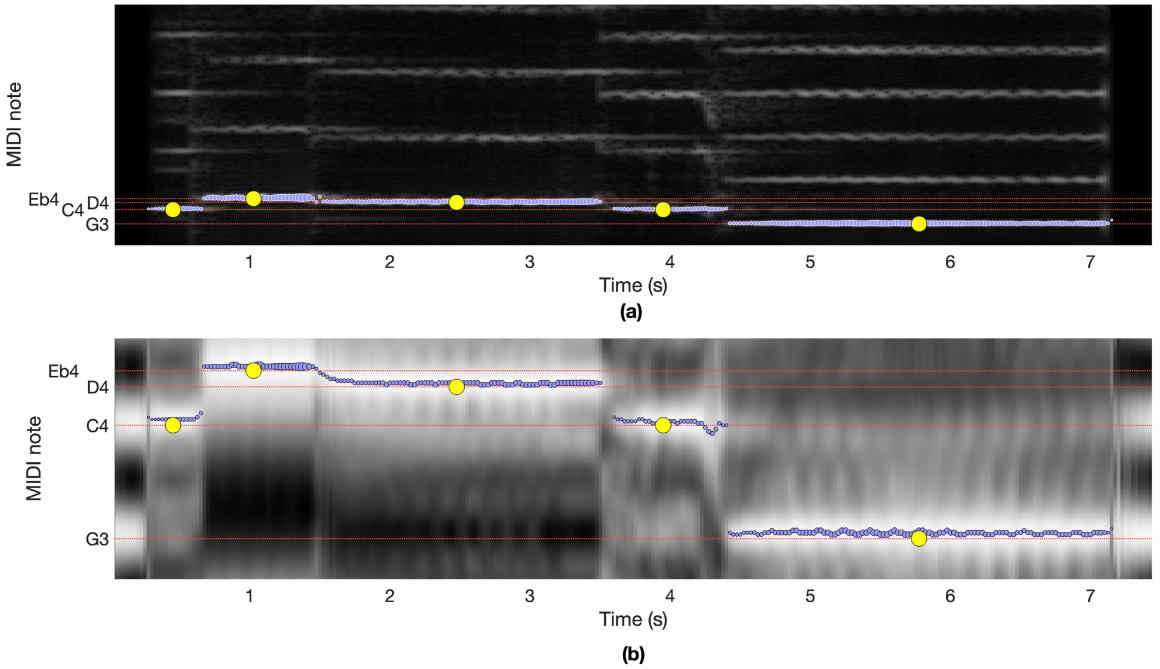


Fig. 5. Cello (sample 6). (a) STFT, (b) STAC. See Supplementary Material[1] SEC. B, for the same figure with synchronized audio. The yellow scatter plots indicate the listener transcriptions shown in Fig. 3; the blue scatter plots indicate the maxima for each frame.

Fig. 8 overlays the annotation results with the sample's STFT. STAC is not included, as it is not suited for pitch detection in inharmonic tones [14]. Pitch perception appears to be primarily driven by spectral modeling, with most perceived pitches corresponding to prominent partials. Although low sensitivity to bass frequencies may explain the absence of B1, it does not account for the perception of B2 in its place. Temporal modeling may explain why a B is perceived, yet it remains unclear why the upper octave is heard rather than the frequency difference itself.

6.5 Power Chord (Sample 13)

6.5.1 Sample Details

The played notes are E2, B2, and E3. The guitar is processed with Native Instrument's Guitar Rig "Rammfire"

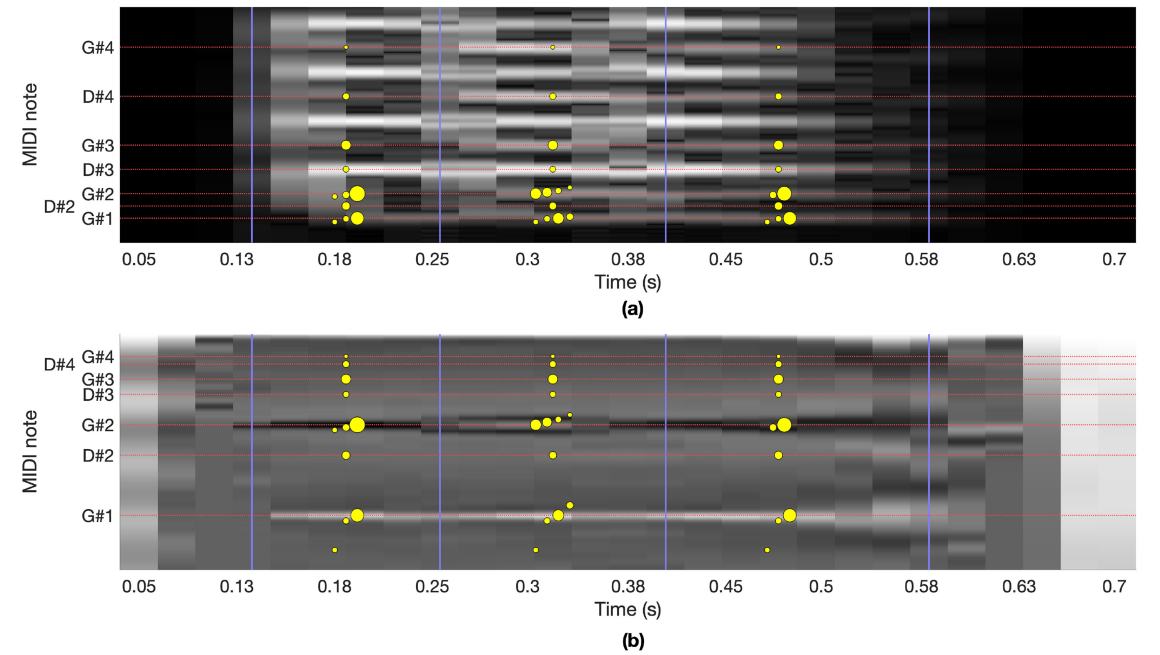


Fig. 6. "Forgiven," bass, 0:00–0:01 (sample 2), weighted audio. See Supplementary Material[1] SEC. C for the same figure with synchronized audio. (a) STFT and (b) STAC. The vertical blue lines indicate note boundaries.

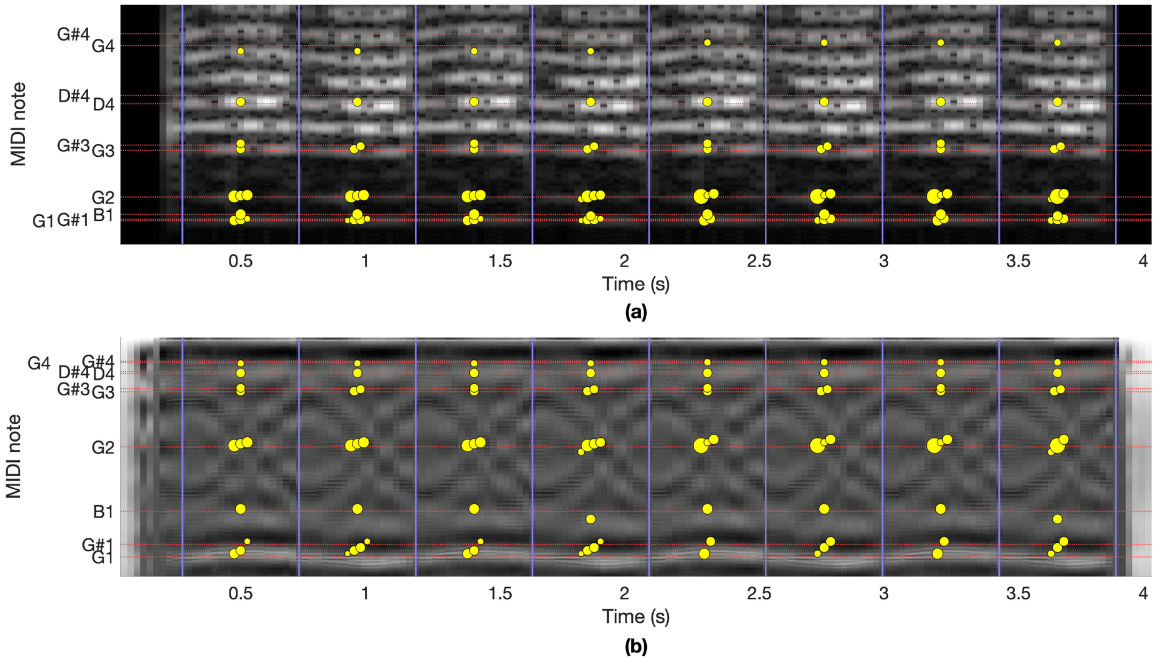


Fig. 7. "Stamina," main synthesizer/bass, 0:40–0:47 (sample 7), weighted audio. See Supplementary Material[1] SEC. D for the same figure with synchronized audio. (a) STFT and (b) STAC. The vertical blue lines show the "note" start.

amp emulation. Distortion applies a nonlinear transformation to the signal, inducing intermodulation distortion. Intermodulation distortion generates partials from each input partial and their combinations [49], resulting in a quasi-harmonic complex tone with E1 as f_0. The partial corresponding to the f_0 is weak.

6.5.2 Perceived Pitches and Signal

Fig. 9(a) suggests that, as in sample 7, the relatively low loudness of the lower harmonics leads to the perception of low pitches (E1, E2) primarily through temporal modeling. Fig. 9(b) further indicates that the preference for E2 may be related to a more clearly defined autocorrelation peak. In contrast, the alignment of higher perceived pitches (D4, E4, A5) with upper partials supports spectral modeling. The variation in perceived pitch resulting from changes in partial energy distribution underscores the role of upper partial loudness in facilitating tonal fission.

6.6 808 Woofer Warfare (Sample 14)

6.6.1 Sample Details

The sample was produced using the Seismic Shock library in Omnisphere. The 808 Woofer Warfare patch, designed by Sonic Extensions, features "tone modes"—presets that shape the spectrum of an 808 kick drum sample. Fig. 10 shows the STFT of the original sample alongside its seven tone modes. Like the original, each mode begins with a continuous frequency drop, with the f_0 settling at C2. Sample 14 uses mode 1, which emphasizes harmonic 5 (E4), a mode favored by the Hyper Music company in production [13].

6.6.2 Perceived Pitches and Signal

As in samples 7 and 13, the STFT, here in Fig. 11(a), suggests an association between higher perceived pitches and spectral modeling, while the low energy of the lower partials supports the perception of low pitches (C1, C2)

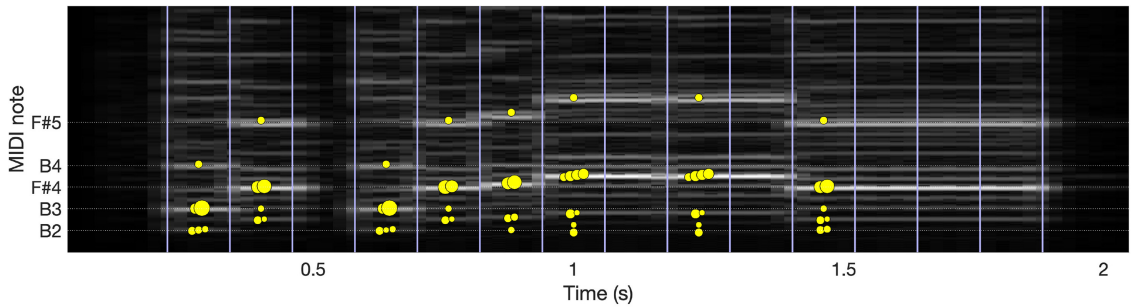


Fig. 8. "No Fun," 0:12–0:14, main synthesizer part (sample 1). See Supplementary Material[1] SEC. E for the same figure with synchronized audio. STFT, weighted audio. The vertical blue lines indicate the positions of the eighth-notes.

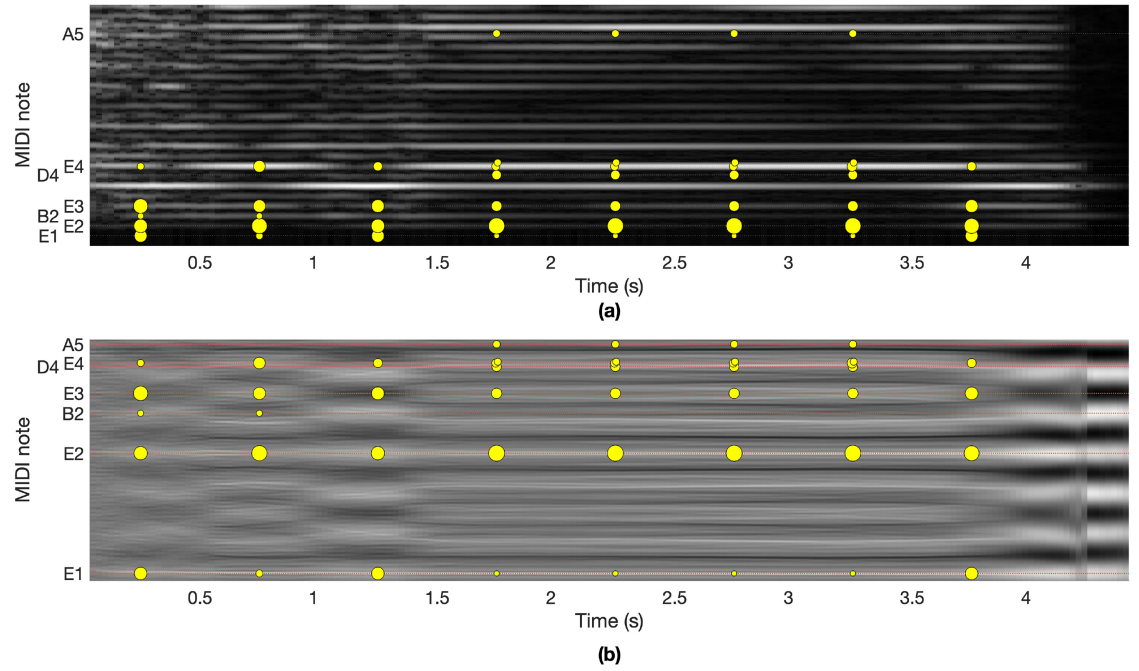


Fig. 9. Power chord (sample 13), weighted audio. See Supplementary Material[1] SEC. F for the same figure with synchronized audio. (a) STFT and (b) STAC.

via temporal modeling. The distinct autocorrelation peak in Fig. 11(b) further reinforces the link between C2 and temporal modeling.

6.7 Signal Analysis: Conclusion

Signal analysis of the Listening Test 2 samples suggests that tonal fission in nonacoustic quasi-harmonic tones (samples 2, 7, 13, and 14) may arise from (1) spectral modeling of loud upper partials and (2) temporal modeling that leads to different pitch outcomes than spectral modeling. Typically, higher-frequency partials are associated with spectral modeling, while lower-frequency partials are linked to temporal modeling.

Similar observations apply to the electronic inharmonic tone (sample 1) but not to the instrumental quasi-harmonic tone (sample 6). The next section proposes that the exceptional status of sample 6 may be due not only to its intrinsic acoustic properties but also to cultural conditioning.

7 INTERPRETATION OF RESULTS THROUGH LISTENER COMMENTS

This section synthesizes feedback from participants in both listening tests, emphasizing interlistener variability, the influence of listening conditions, and the fluid boundary between pitch and timbre. The comments indicate that pitch perception is a context-dependent, culturally mediated, and individually variable construct shaped by musical background, listening conditions, and perceptual focus, rather than an intrinsic property of sound. They further support the hypothesis that pitch perception can, in some cases, be multistable.

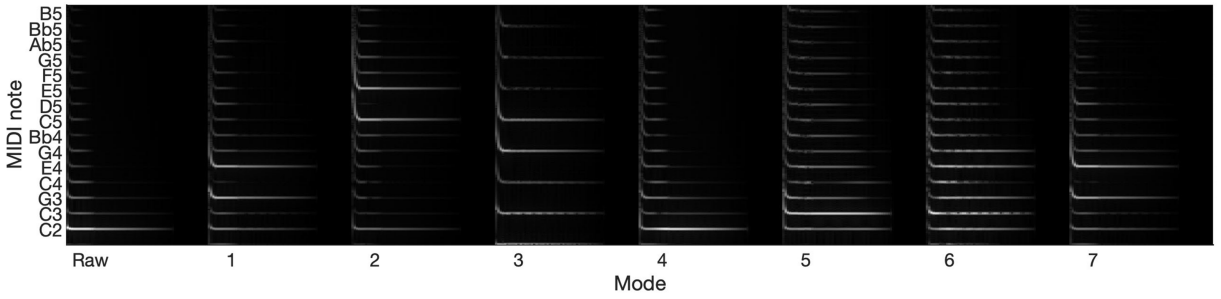


Fig. 10. Omnisphere, Seismic Shock library, 808 Woofer Warfare patch, STFT for the seven modes. Unweighted audio. See Supplementary Material[1] SEC. G for the same figure with synchronized audio.

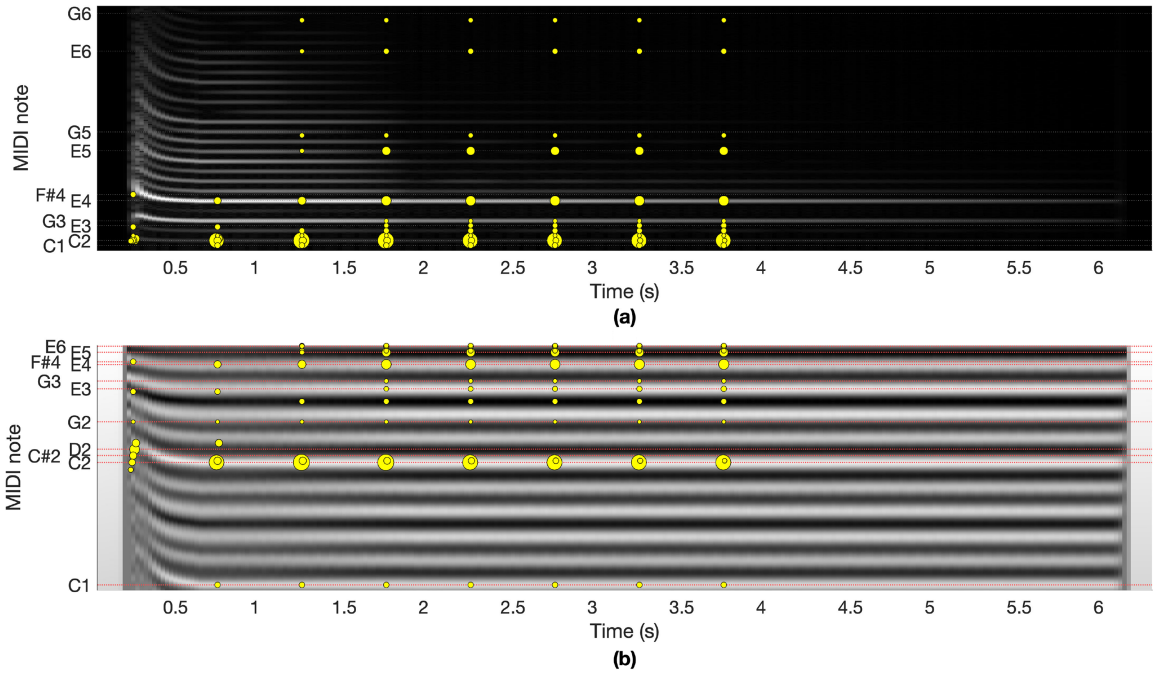


Fig. 11. 808 Woofer Warfare, mode 1 (sample 14), weighted audio. See Supplementary Material[1] SEC. G for the same figure with synchronized audio. (a) STFT and (b) STAC.

The interpretation begins with Listening Test 2, as the expert comments better help clarify the notions involved. The nonexpert comments (Test 1) are then categorized by comparison.

7.1 Comments From Expert Listeners (Listening Test 2)

Expert listeners (ELs) had the following backgrounds: EL 1, 2, 4, and 6 were familiar with both Western classical music (WCM) and CPM; EL 3 had a background in CPM; EL 5 had a background in WCM and psychoacoustics; and EL 7 had a background in WCM.

- Interlistener variability: ELs diverged in how they interpreted the tones. EL 6 and 7, both with classical backgrounds, tended to associate each tone with a single pitch, sometimes relegating additional pitches to the realm of timbral texture. In contrast, EL 3, experienced in rap production, regarded multipitch perception as a normative aspect of tone design. In the case of acoustic tones, EL 4 noted the distinction between reporting what is actually heard (multiple pitches) and what is inferred to have been played (a single note).
- Effect of listening conditions: Perceived pitch varied notably with playback conditions. EL 1–4 and 6 reported changes in pitch perception depending on the monitoring system, listening position, and playback level. These effects align with known variations in frequency response across speaker types [50], listener angles [51], and loudness curves [35]. While pitch perception in acoustic samples (cello) remained stable, synthetic tones were sensitive to these variables.
- Perceptual flexibility: Three listeners noted that, even under fixed listening conditions, repeated exposure allowed them to shift attention between different pitches—effectively choosing which ones to perceive within a limited set. This perceptual flexibility is further discussed in SEC. 8.2.
- Pitch–timbre ambiguity: All ELs found it difficult to disentangle pitch from timbre in nonacoustic tones. In some cases, even though pitches were perceived, they were interpreted as timbral features instead of distinct melodic elements. EL 7, for example, reported hearing multiple pitches but chose to interpret them as part of the tone's character rather than as discrete entities.
- Pitch identification in the lower register: Several listeners, notably EL 4 and 7, highlighted challenges in identifying precise pitch values in the lower frequency range (MIDI octaves 1 and 2). They attributed this difficulty both to the perceptual characteristics of low tones—specifically, their reduced pitch strength, as noted by Huron [52, p. 8]—and to limited exposure to such registers in classical and earlier popular music [45, 53].
- Octave ambiguity in the lower register: EL 3 remarked that in rap production, "bass" sounds are not expected to convey a specific octave; a single keypress may trigger energy across two octaves, consistent with prior observations [13]. EL 5 and 7 reported that octaves in the lower register often merged perceptually into a single stream in the sense of Breg-

man [22]. EL 5 compared this to the effect of parallel octaves in Romantic piano music.
- Effects of non–12-TET tuning: EL 1 and 2 observed that deviations from 12-tone equal temperament, as documented in Vitalic's music [14], complicated the task of assigning pitch names. They remarked that such deviations made it more difficult to identify independent lines.

7.2 Comments From Nonexpert Listeners (Listening Test 1)

The nonexpert listeners were students in an engineering program at a university.

- Interlistener variability: Unlike experts, nonexpert listeners described most Vitalic samples in consistent terms. They typically identified distinct "low" and "high" layers and considered high-frequency buzzing components (e.g., in samples 1, 7, 9, and 12) to be melodic. They found it difficult to count the number of simultaneous pitches during rapid changes and noted that masking between layers made this task harder.
- Perception of low-frequency melodies: Unlike experts, who often struggled to assign clear pitches in the low register, two nonexpert listeners reported that low-frequency melodic lines were easier to follow than other lines, as they were well separated from other elements.
- Perception of upper harmonics in acoustic sources: Two nonexpert participants perceived a second melodic layer in the cello sample, attributing it to overtone richness. That only nonexperts mentioned this perception runs counter to Mersenne's historical claim that only the "best ears" can detect upper harmonics [24].
- Violin reverberation and perceived simultaneity: In the violin sample (sample 10), reverberation led one nonexpert to report two simultaneous pitches, while others hesitated. This reflects a gap between the number of acoustically present tones and what is perceptually interpreted as distinct pitches. Indeed, EL 4 noted that while multiple pitches were always present, they were perceived as simultaneous only in short (subsecond) isolated extracts. Refer to Supplementary Material[1] SEC. H for a video showing that multiple tones are indeed physically present due to reverberation, even if not always perceived as simultaneous.

7.3 Interpretation of Listener Comments

The listener comments suggest a fragmented and context-dependent landscape of pitch perception.

- Experts varied in their pitch interpretation strategies, often influenced by musical background. Those trained in classical music tended toward single-pitch reporting, while others embraced tonal fission and ambiguity as stylistically appropriate.
- Pitch perception was especially unstable in the lower register, where octave ambiguity and tonal fusion frequently obscured register identification.
- Listening conditions—system, position, and level—significantly impacted pitch perception in synthetic tones.
- Several ELs demonstrated perceptual flexibility, suggesting that pitch attention can be selectively directed, not just passively received.
- The boundary between pitch and timbre was blurred in complex tones, often preventing clear categorization.
- Nonexpert listeners tended to report more consistent perceptions across samples. They identified clear high and low layers, on some occasions showing sensitivity to melodic content that experts overlooked.
- The perception of upper harmonics in monophonic instruments by nonexperts challenges assumptions about auditory expertise.

Taken together, these findings confirm that pitch should not be treated as an objective, intrinsic property of sound, but as a perceptual construct—shaped by listening context, cultural familiarity, and individual focus. SEC. 8.2 elaborates on the multistable aspect of pitch, while SEC. 8.1 discusses implications for theoretical models.

8 DISCUSSION

8.1 Rethinking Pitch in Perception and Music Analysis

WCM is grounded in notated scores that encode properties such as pitch. Instruments in this tradition typically produce harmonic or quasi-harmonic tones via acoustic resonance, with the perceived pitch usually aligning with the fundamental frequency (f_0). This framework has shaped key tasks in music information retrieval, such as pitch tracking [54, 55] and audio-to-MIDI alignment, where MIDI files serve as symbolic representations [56, 57]. The reliance on notation in analysis has led to what Middleton [58] calls the "reification of the score," treating it as an idealized version of the music.

This model assumes that pitch is an ontological property of a tone—that each tone corresponds to a single, context-independent and listener-independent pitch. However, such assumptions may be challenged. Prior research [13, 14], this study, and the producer's deliberate use of pitch uncertainty suggest that these views may be inadequate for musical contexts beyond WCM.

Electronic music producers like Vitalic and the Hyper Music producers [13] deliberately manipulate pitch perception by varying the number of pitches per tone and introducing pitch ambiguity as a compositional parameter. The commercial success of their work points to an alternative paradigm, in which tones may be designed to transmit not fixed pitches but sets of possible ones. The use of

power chords in rock and metal further indicates that this approach may extend beyond electronic music into other popular genres.

This shift prompts a redefinition of what constitutes a musical tone. Rather than equating a tone with a single perceived pitch [59], an alternative view defines it as a perceptual stream of partials that share a common fate [30, 60], with some pitches being more likely to be perceived than others. This definition is adopted in SEC. 2.3 and in [14]. From this perspective, pitch may be described not as a fixed value but as a probability distribution. A related interpretation treats the probability of perceiving a given pitch as a measure of pitch strength [61].

This study has focused exclusively on isolated tone sequences. Tones presented in a musical context—beyond the scope of this work—may behave differently. A preliminary listening test in [13] suggests that the perceived pitch of a low pitch-strength tone changes when played alongside another track. Vitalic also noted, in a public interview [62], that the pitches perceived from tones similar to those used here depend on the surrounding context.

8.2 The Musical Role of Multistable Perception

Multistability arises when a single physical stimulus produces alternating subjective percepts [63]. A classic example is the "Necker cube" [64, p. 336], a two-dimensional drawing that can be seen with either the lower-left or upper-right square as the front face. In audio, multistability (in this case, bistability) includes: (1) the tritone paradox, where a pair of Shepard tones [2] separated by a tritone is heard as ascending or descending [65], and (2) an A-B tone sequence heard as one stream (ABA-ABA) or two: A-A-A-A and -B-B-B- [66].

SEC. 7.1 notes that listeners in test 2 could deliberately choose one pitch set over another, indicating multistable perception—suggesting quasi-harmonic complex tones can yield multistable percepts. Deliberately constructing tones as multistable may help sustain listener interest. Madison and Schiölde [67] found that musical appreciation declines with repeated exposure in a single session. A relevant pattern appears in Vitalic's tracks, which often feature heavy repetition (see Supplementary Material[1] SEC. C). Orbach et al. [68] explain Necker cube reversals via "satiation of orientation": once a perceptual process is saturated, a switch occurs. This suggests that producers may use multistable tones to enable perceptual switching, delaying satiation and maintaining appeal across repeated hearings. These hypotheses warrant further study.

9 CONCLUSION

The electronic music producer Vitalic deliberately designs quasi-harmonic tones that convey multiple simultaneous pitches—a phenomenon referred to here as *tonal fission*. Building on prior studies that examined the signal properties of such tones in the music of Vitalic and others [13, 14], the present study employed listening tests to further explore this perceptual effect.

Listening Test 1 investigated the number of simultaneous pitches perceived in electronic and acoustic tones within a musical context. Results indicate that quasi-harmonic tones—particularly electronic ones—can convey multiple pitches simultaneously, with perceptual outcomes varying across listeners.

In Listening Test 2 (SEC. 5), musically trained participants transcribed the pitches they perceived in tone sequences. These transcriptions confirmed that, apart from the acoustic tone sequence, multiple pitches were often perceived simultaneously, with the specific identified pitches varying across participants. Signal analysis suggests that this multiplicity of perceived pitches (tonal fission) may result from spectral modeling of prominent partials, potentially conflicting with the outcomes of temporal modeling.

SEC. 7 presented listener comments that emphasized interlistener variability, the influence of listening conditions, and the ambiguity between pitch and timbre perception. These findings underscore the contextual and subjective nature of pitch perception in electronically produced music and suggest that tones may serve different functions in popular music—transmitting a set of pitches that may or may not be perceived—than in Western classical traditions, where (typically) a single intended pitch is expected to be perceived corresponding to each note.

SEC. 8 explored broader implications, including challenges to conventional definitions of musical tones and to pitch-tracking methods in music analysis. It also addressed the relevance of multistable perception in quasi-harmonic tones.

In electronic music, pitch may not be an intrinsic property of a sound object, but rather a listener-constructed, context-dependent percept. A single harmonic tone in a musical context can be deliberately designed to project multiple, ambiguous pitches to listeners through tonal fission. This challenges the definition of a "note" in electronic music and perhaps more broadly in CPM.

These conclusions open several directions for future research: for instance, better examining the cognitive mechanisms underlying tonal fission; exploring its pervasiveness in a broader range of musical contexts; further investigating the potential transition from active to resultant pitch in CPM; and confirming or refuting the existence of multistable pitch perception in this context.

10 ACKNOWLEDGMENT

Many thanks go to Cyran Aouameur, Amaury Delort, Matthias Demoucron, and Stefan Lattner, Sony Computer Science Laboratories, Paris, France; Michael Turbot, Sony AI, Zürich, Switzerland; and Yann Macé and Luc Leroy from the music production company Hyper Music.

11 REFERENCES

[1] 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://doi.org/10.1017/S0007087400015089.

[2] R. N. Shepard, "Circularity in Judgments of Relative Pitch," *J. Acoust. Soc. Am.*, vol. 36, no. 12, pp. 2346–2353 (1964 Dec.). https://doi.org/10.1121/1.1919362.

[3] B. C. J. Moore, B. R. Glasberg, H. J. Flanagan, and J. Adams, "Frequency Discrimination of Complex Tones; Assessing the Role of Component Resolvability and Temporal Fine Structure," *J. Acoust. Soc. Am.*, vol. 119, no. 1, pp. 480–490 (2006 Jan.). https://doi.org/10.1121/1.2139070.

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

[5] R. Plomp, *Aspects of Tone Sensation: A Psychophysical Study* (Academic Press, New York. NY, 1976).

[6] W. Dixon Ward, "Musical Perception," in J. Tobias (Ed.), *Foundations of Modern Auditory Theory*, vol. 1, pp. 405–447 (Academic Press, New York, NY, 1970). https://doi.org/10.1016/B978-0-12-691901-1.X5001-3.

[7] G. Bloothooft, E. Bringmann, M. Van Cappellen, J. B. Van Luipen, and K. P. Thomassen, "Acoustics and Perception of Overtone Singing," *J. Acoust. Soc. Am.*, vol. 92, no. 4, pp. 1827–1836 (1992 Oct.). https://doi.org/10.1121/1.403839.

[8] C. Fox and E. Fallowfield, "Editorial: The Art of the String Multiphonic," *Tempo*, vol. 74, no. 291, pp. 3–5 (2020 Jan.). https://doi.org/10.1017/S0040298219000949.

[9] E. Fallowfield, "Cello Multiphonics: Technical and Musical Parameters," *Tempo*, vol. 74, no. 291, pp. 51–69 (2020 Jan.). https://doi.org/10.1017/S0040298219000974.

[10] E. Deruty, M. Grachten, S. Lattner, J. Nistal, and 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. 35–49 (2022 Feb.). https://doi.org/10.5334/tismir.100.

[11] E. Deruty and M. Grachten, "'Melatonin': A Case Study on AI-Induced Musical Style," presented at the *3rd Conf. on AI Music Creativity* (Online) (2022 Sep.). https://doi.org/10.5281/zenodo.7088302.

[12] E. Deruty and M. Grachten, "Insights on Harmonic Tones From a Generative Music Experiment," presented at the *15th Intl. Workshop on Machine Learning and Music* (Vilnius, Lithuania) (2024 Sep.).

[13] E. Deruty, L. Leroy, Y. Macé, and D. Meredith, "Methods for Pitch Analysis in Contemporary Popular Music: Highlighting Pitch Uncertainty in Primaal's Commercial Works," *Journal of New Music Research*, in press (2025). https://arxiv.org/abs/2502.08131.

[14] 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.

[15] D. Huron and J. Berec, "Characterizing Idiomatic Organization in Music: A Theory and Case Study of Musical Affordances," *Empir. Musicol. Rev.*, vol. 4, no. 1, pp. 103–122 (2009 Jul.). https://doi.org/10.18061/1811/44531.

[16] A. Riou, B. Torres, B. Hayes, et al., "PESTO: Real-Time Pitch Estimation With Self-Supervised Transposition-Equivariant Objective," *Trans. Int. Soc. Music Inf. Retr.*, vol. 8, no. 1, pp. 334–352 (2025 Sep.). https://doi.org/10.5334/tismir.251.

[17] D. E. Berlyne, *Conflict, Arousal, and Curiosity* (McGraw-Hill, New York, NY, 1960).

[18] MusicRadar Tech, "Vitalic in the Studio With Future Music," (2013 Oct.). https://youtu.be/ogjeXBY-iEI?t=180 (accessed 8 Apr. 2026).

[19] C. Stumpf, *Tonpsychologie*, vol. 2 (S. Hirzel, Leipzig, Germany, 1890).

[20] L. A. DeWitt and R. G. Crowder, "Tonal Fusion of Consonant Musical Intervals: The Oomph in Stumpf," *Percept. Psychophys.*, vol. 41, no. 1, pp. 73–84 (1987 Jan.). https://doi.org/10.3758/BF03208216.

[21] L. P. A. S. van Noorden, *Temporal Coherence in the Perception of Tone Sequences*, Ph.D. thesis, Technische Hogeschool Eindhoven, Eindhoven, the Netherlands (1975 Feb.). https://pure.tue.nl/ws/files/3389175/152538.pdf.

[22] A. S. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge, MA, 1994). https://mitpress.mit.edu/9780262521956/.

[23] B. C. J. Moore and H. E. Gockel, "Properties of Auditory Stream Formation," *Philos. Trans. R. Soc. B Biol. Sci.*, vol. 367, no. 1591, pp. 919–931 (2012 Apr.). https://doi.org/10.1098/rstb.2011.0355.

[24] M. Mersenne, *Harmonie Universelle* (Chez Sébastien Cramoisy, Paris, France, 1636).

[25] J. P. Rameau, *Démonstration du Principe de L'Harmonie, Servant de Base à Tout l'Art Musical Théorique et Pratique* (Durand et Pissot, Paris, France, 1750). https://gallica.bnf.fr/ark:/12148/bpt6k1082246/.

[26] H. L. F. von Helmholtz, *On the Sensations of Tone as a Physiological Basis for the Theory of Music*, A. J. Ellis (Trans.) (Longmans, Green, London, UK, 1885).

[27] 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.

[28] R. Plomp, *The Intelligent Ear: On the Nature of Sound Perception* (Psychology Press, New York, NY, 2001). https://doi.org/10.4324/9781410604255.

[29] E. Deruty, "Inharmonic Tones and Listener Perception in Vitalic's 'No Fun,'" (2025 Oct.). https://medium.com/@derutycsl/inharmonic-tones-and-listener-perception-in-vitalics-no-fun-978965a355d3 (accessed 9 Apr. 2026).

[30] R. Rasch and R. Plomp, "The Perception of Musical Tones," in D. Deutsch (Ed.), *The Psychology of Music*, pp. 89–112 (Academic Press, San Diego, CA, 1999), 2nd ed. https://doi.org/10.1016/B978-012213564-4/50005-6.

[31] 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.

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

[33] L. Rabiner, M. Cheng, A. Rosenberg, and 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.

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

[35] H. Fletcher and 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.

[36] D. W. Robinson and R. S. Dadson, "A Re-Determination of the Equal-Loudness Relations for Pure Tones," *Br. J. Appl. Phys.*, vol. 7, no. 5, pp. 166–181 (1956 May). https://doi.org/10.1088/0508-3443/7/5/302.

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

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

[39] B. O. Olusanya, A. C. Davis, and H. J. Hoffman, "Hearing Loss Grades and the International Classification of Functioning, Disability and Health," *Bull. World Health Organ.*, vol. 97, no. 10, pp. 725–728 (2019 Sep.). https://doi.org/10.2471/BLT.19.230367.

[40] L. F. Aluas, *The "Quatour Pincipalia Musicae": A Critical Edition and Translation, With Introduction and Commentary*, Ph.D. thesis, Indiana University, Bloomington, IN (1996 May). https://www.proquest.com/openview/251d1cbafdc43d407c5a5557263cd619.

[41] J. C. Nunes, A. Ordanini, and F. Valsesia, "The Power of Repetition: Repetitive Lyrics in a Song Increase Processing Fluency and Drive Market Success," *J. Consum. Psychol.*, vol. 25, no. 2, pp. 187–199 (2015 Apr.). https://doi.org/10.1016/j.jcps.2014.12.004.

[42] B. C. J. Moore, B. R. Glasberg, and 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.

[43] T. M. Shackleton and R. P. Carlyon, "The Role of Resolved and Unresolved Harmonics in Pitch Perception and Frequency Modulation Discrimination," *J. Acoust. Soc. Am.*, vol. 95, no. 6, pp. 3529–3540 (1994 Jun.). https://doi.org/10.1121/1.409970.

[44] L. Richter, "Ptolemy [Klaudios Ptolemaios; Claudius Ptolemaeus]," *Grove Music Online*, paper 22510 (2001 Jan.). https://doi.org/10.1093/gmo/9781561592630.article.22510.

[45] E. Deruty, "Harmonic and Transposition Constraints Arising From the Use of the Roland TR-808 Bass Drum," in Proc. *25th Intl. Society for Music Information Retrieval Conf.*, pp. 78–85 (San Francisco, CA) (2024 Nov.). https://zenodo.org/records/14877286.

[46] M. Winter, R. Wannamaker, L. Pratt, L. Polansky, and J. Tenney, *From Scratch: Writings in Music Theory* (University of Illinois Press, Champaign, IL, 2014). https://muse.jhu.edu/pub/34/monograph/book/42398.

[47] 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.

[48] A. de Cheveigné and H. Kawahara, "YIN, a Fundamental Frequency Estimator for Speech and Music," *J. Acoust. Soc. Am.*, vol. 111, no. 4, pp. 1917–1930 (2002 Apr.). https://doi.org/10.1121/1.1458024.

[49] P. Newell, *Recording Studio Design* (Taylor & Francis, New York, NY, 2017). https://doi.org/10.4324/9781315675367.

[50] P. R. Newell, K. R. Holland, and J. P. Newell, "The Yamaha NS10M: Twenty Years a Reference Monitor. Why?" in Proc. *Institute of Acoustics*, vol. 23, Part 8, pp. 29–40 (Stratford, UK) (2001 Nov.).

[51] W. Evans, J. Dyreby, S. Bech, S. Zielinski, and F. Rumsey, "Effects of Loudspeaker Directivity on Perceived Sound Quality-A Review of Existing Studies," presented at the *126th Conv. of the Audio Engineering Society* (2009 May), paper 7745. https://aes2.org/publications/elibrary-page/?id=14941.

[52] D. Huron, "Tone and Voice: A Derivation of the Rules of Voice-Leading From Perceptual Principles," *Music Percept.*, vol. 19, no. 1, pp. 1–64 (2001 Fall). https://www.jstor.org/stable/pdf/10.1525/mp.2001.19.1.1.pdf.

[53] M. J. Hove, P. Vuust, and J. Stupacher, "Increased Levels of Bass in Popular Music Recordings 1955–2016 and Their Relation to Loudness," *J. Acoust. Soc. Am.*, vol. 145, no. 4, pp. 2247–2253 (2019 Apr.). https://doi.org/10.1121/1.5097587.

[54] T. Drugman, G. Huybrechts, V. Klimkov, and A. Moinet, "Traditional Machine Learning for Pitch Detection," *IEEE Signal Process. Lett.*, vol. 25, no. 11, pp. 1745–1749 (2018 Nov.). https://doi.org/10.1109/LSP.2018.2874155.

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

[56] S. Ewert, M. Muller, V. Konz, D. Mullensiefen, and G. A. Wiggins, "Towards Cross-Version Harmonic Analysis of Music," *IEEE Trans. Multimed.*, vol. 14, no. 3, pp. 770–782 (2012 Jun.). https://doi.org/10.1109/TMM.2012.2190047.

[57] E. Benetos, S. Dixon, Z. Duan, and S. Ewert, "Automatic Music Transcription: An Overview," *IEEE Signal Process. Mag.*, vol. 36, no. 1, pp. 20–30 (2019 Jan.). https://doi.org/10.1109/MSP.2018.2869928.

[58] R. Middleton, *Studying Popular Music* (Open University Press, Buckingham, UK, 1990).

[59] B. C. J. Moore, B. R. Glasberg, and 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.

[60] 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, London, UK, 1938). https://doi.org/10.1037/11496-005.

[61] E. Deruty, "An Introduction to Pitch Strength in Contemporary Popular Music Analysis and Production," in *Innovation in Music: Current Research Perspectives*, pp. 77–93 (Focal Press, London, UK, 2026). https://doi.org/10.4324/9781003475675-8.

[62] Sony CSL (Paris)Vitalic, "VITALIC x Sony CSL - Paris 'It Has Been Therapy Sessions!'," 2026 Apr.). https://youtu.be/BDqfk_RZHcY (accessed 16 Apr. 2026).

[63] J.-L. Schwartz, N. Grimault, J.-M. Hupé, B. C. J. Moore, and D. Pressnitzer, "Multistability in Perception: Binding Sensory Modalities, an Overview," *Philos. Trans. R. Soc. B Biol. Sci.*, vol. 367, no. 1591, pp. 896–905 (2012 Apr.). https://doi.org/10.1098/rstb.2011.0254.

[64] L. A. Necker, "LXI. Observations on Some Remarkable Optical Phænomena Seen in Switzerland; and on an Optical Phænomenon Which Occurs on Viewing a Figure of a Crystal or Geometrical Solid," *Lond. Edinb. Dubl. Philos. Mag. J. Sci.*, vol. 1, no. 5, pp. 329–337 (1832 Nov.). https://doi.org/10.1080/14786443208647909.

[65] D. Deutsch, "The Tritone Paradox: An Influence of Language on Music Perception," *Music Percept.*, vol. 8, no. 4, pp. 335–347 (1991 Jul.). https://doi.org/10.2307/40285517.

[66] D. Pressnitzer and J.-M. Hupé, "Temporal Dynamics of Auditory and Visual Bistability Reveal Common Principles of Perceptual Organization," *Current Biol.*, vol. 16, no. 13, pp. 1351–1357 (2006 Jul.). https://doi.org/10.1016/j.cub.2006.05.054.

[67] G. Madison and G. Schiölde, "Repeated Listening Increases the Liking for Music Regardless of Its Complexity: Implications for the Appreciation and Aesthetics of Music," *Front. Neurosci.*, vol. 11, paper 147 (2017 Mar.). https://doi.org/10.3389/fnins.2017.00147.

[68] J. Orbach, D. Ehrlich, and H. A. Heath, "Reversibility of the Necker Cube: I. An Examination of the Concept of 'Satiation of Orientation,'" *Percept. Motor Skills*, vol. 17, no. 2, pp. 439–458 (1963 Oct.). https://doi.org/10.2466/pms.1963.17.2.439.

[69] Vitalic, "No Fun," in *OK Cowboy* (Different Records / PIAS, London, UK, 2005). https://youtu.be/f8K7Pbxhgps.

[70] Vitalic, "Forgiven," in *Confess* (Citizen Records / Virgin Music France, Paris, France, 2023). https://youtu.be/j5Z-YCwcMts.

[71] Vitalic, "La Mort sur le Dance Floor," in *Rave Age* (Different Records / PIAS, London, UK, 2012). https://youtu.be/jhsIyAVz1eI.

[72] Vitalic, "Sweet Cigarette," in *Voyager* (Clivage Music / Universal, Dijon / Paris, France 2017). https://youtu.be/2ljrY75C_D8.

[73] B. Tchaikovsky, "Solo Cello Suite in D Minor: III. Aria. Adagio," performed by Christopher Marwood (Naxos, London, UK, 2018). https://youtu.be/KEBcrzhxjTQ.

[74] Vitalic, "Stamina," in *Rave Age* (Different Records / PIAS, London, UK, 2012). https://youtu.be/OebhxAJnF-E.

[75] Vitalic, "Next I'm Ready," in *Rave Age* (Different Records / PIAS, London, UK, 2012). https://youtu.be/FhOfNgDNia8.

[76] J. S. Bach, "Partita II BWV 1004 in D Minor: V. Ciaccona," performed by Isabelle Faust (Harmonia Mundi, Arles, France, 2010).

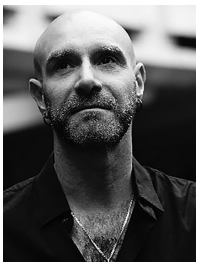
[77] Vitalic, "See the Sea (Blue)," in *Flashmob* (Different Records / PIAS, London, UK, 2009). https://youtu.be/Q7KpYxb4Xhg.

THE AUTHORS


Emmanuel Deruty


David Meredith


Maarten Grachten


Pascal Arbez-Nicolas


Andreas Hasselholt Jørgensen


Oliver Søndermølle Hansen


Christian Nørkær Petersen


Magnus Stensli

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) in Paris and Tokyo, focusing on AI-assisted music production.

•

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

•

Maarten Grachten is a consultant in applied machine learning in both academic and industrial settings. His work spans grant acquisition, project management, academic writing, and software development. Key application areas include music information retrieval, music technology, audio signal processing, and natural language processing.

•

Pascal Arbez-Nicolas, known professionally as Vitalic, is a French electronic music producer who blends techno, electro, and house with classical, rock, and ’70s electronic music. His career reflects a continuous evolution of electronic music styles while maintaining a rebellious and experimental edge.

•

Andreas Hasselholt Jørgensen is a master’s student in medialogy at Aalborg University (AAU), specializing in the design and evaluation of human-centered interactive media technologies.

•

Oliver Søndermølle Hansen is a master’s student in medialogy at Aalborg University (AAU), specializing in the design and evaluation of human-centered interactive media technologies.

•

Christian Nørkær Petersen is a master’s student in medialogy at Aalborg University (AAU), specializing in the design and evaluation of human-centered interactive media technologies.

•

Magnus Stensli is a master’s student in artificial intelligence, vision, and sound at Aalborg University (AAU) with a bachelor’s degree in medialogy.