MEASURING SENSORY DISSONANCE IN MULTI-TRACK MUSIC RECORDINGS: A CASE STUDY WITH WIND QUARTETS

Simon Schwär **Stefan Balke** **Meinard Müller**
International Audio Laboratories Erlangen, Germany
{simon.schwaer, stefan.balke, meinard.mueller}@audiolabs-erlangen.de

ABSTRACT

Sensory dissonance (SD) quantifies the interference between partials in a mixture of simultaneously sounding tones and correlates with the perceived dissonance or unpleasantness of this mixture. While it is mainly studied in music perception, often using synthetic signals or symbolic inputs, in this paper, we focus on a practical application and investigate SD as a tool for analyzing the interactions between voices in multi-track music recordings. Using visualization and statistical analysis on an existing dataset of four-part chorales recorded with various wind instruments, we examine how timbre, tuning, and score influences SD. To do this, we introduce the notion of relative SD, which quantifies how individual voices in a multi-track recording contribute to overall SD of their polyphonic mixture. In addition to discussing practical aspects of measuring SD between and within real music signals, our case study shows potential benefits and limitations of using SD as an analysis tool in music production, for example, to inform or automate tasks like take selection or equalization.

1. INTRODUCTION

In music production, the creative process of editing and mixing can often be aided by objective measures of sound properties, such as displaying loudness differences with a *level meter* or visualizing phase differences between stereo channels with a *goniometer*. To our knowledge, a property that has not yet been considered in this context is the dissonance in a track or recording. With the interplay between consonance and dissonance being considered a core component of musical expression [1], such a measure could give insights into musical properties of a mix, both for relative comparisons (e.g., to evaluate intonation and voice blending between different tracks) and as an absolute quantity (e.g., for retrieving sections with high dissonance).

The musical concept of dissonance is a multi-faceted issue [3] with strong cultural influences [4]. While acoustically measurable effects [5, 6] have been found to correlate with subjective dissonance ratings in isolated intervals and chords [7, 8], they can only serve as indirect proxies for musical dissonance. In this paper, we explore *sensory dissonance* (SD) [9], a measure for the interactions between tonal components in a complex sound [10, 11], as a way to quantify dissonance in recorded music performances directly from audio. Similar to the goniometer that does not measure the subjective impression of the stereo image, the goal is not to analyze perceptual properties of SD, but to better understand its behavior in a realistic musical scenario, exploring what it can reveal about the relations between individual voices in a multi-track recording.

We approach this question with an exploratory case study using the ChoraleBricks dataset [2]. It comprises

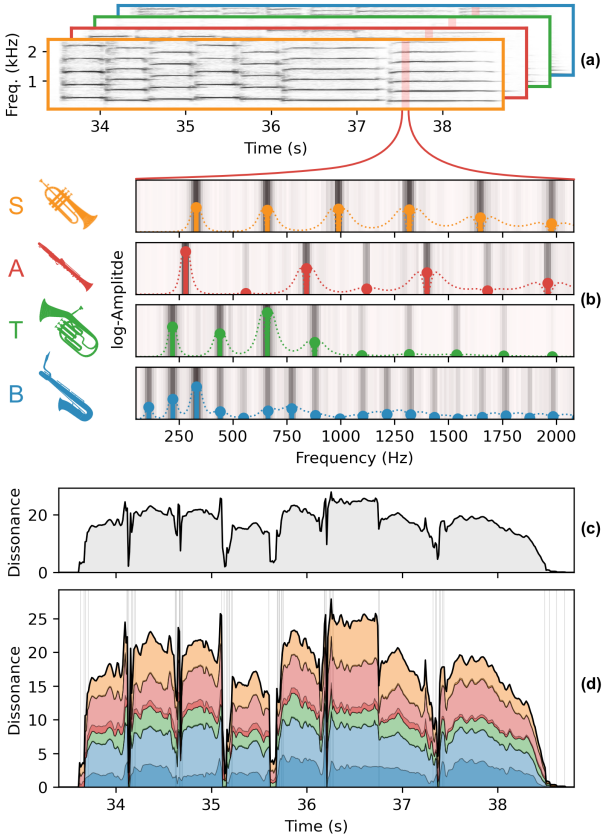

Figure 1. Excerpt from the chorale `GE1` in ChoraleBricks [2]. **(a)** Spectrograms of individual voices played with trumpet (S), clarinet (A), baritone (T) and baritone sax (B). **(b)** Peak representation $\mathcal{P}_v(m)$ for frame m and each voice v. Dotted lines illustrate the amplitude-weighted dissonance kernels. **(c)** Overall dissonance D of the excerpt. **(d)** Relative SD by voice (light: $D_{v,\overline{v}}$, dark: $D_{v,v}$).

performances of ten Baroque chorales, each with four voices—soprano (S), alto (A), tenor (T), and bass (B)—that are recorded in isolation and played on several different wind instruments. This way, the dataset provides a controlled scenario for discerning the influence of different musical aspects, including timbre (instrument characteristics), tuning (pitch deviations), and score (chords and voicings). As an initial example, we consider an excerpt from the chorale "Befiehl Du Deine Wege" (GE1 in ChoraleBricks) in Fig. 1, played with trumpet (S, orange), clarinet (A, red), baritone horn (T, green), and baritone saxophone (B, blue). After estimating tonal components (in this case, the harmonics) over time from each of the four tracks (Fig. 1a and b), we can calculate the overall SD (shown in Fig. 1c) and split it into the relative contributions of each voice (shown in Fig. 1d). This visualization—inspired by Sethares' *dissonance score* [9]—reveals several aspects about SD. For instance, we can observe differences between chords (Fig. 1d shows the note onsets of each voice as vertical lines for orientation), variations in the contributions of individual voices, and local fluctuations within chords (e.g., around 38 seconds). Exploring these effects in detail is a main objective of this paper.

In this context, we make three main contributions. First, we formalize the notion of *relative SD* (Section 2.1), measuring the contribution of individual tracks to the overall SD in a music performance, and consider practical aspects of calculating SD from real signals (Section 2.2). Second, we examine the influence of timbre, tuning, and score on relative SD by introducing new visualizations and conducting systematic experiments with ChoraleBricks (Section 3). Third, we outline possible applications of SD for tasks in music production, including take selection and equalization (Section 4), and discuss limitations and ambiguities that arise when measuring SD. A Python library with all tools used in the experiments and our new chord annotations for ChoraleBricks are available online. [1]

2. SENSORY DISSONANCE

The concept of sensory dissonance (SD) plays a significant role in music research, where it has been employed in perceptual models [8, 12–14], for the bottom-up construction of scales and music theories [9, 15], and to analyze intonation and tuning [16, 17]. In these contexts, several models have been proposed to quantify SD [8, 12–15], all based on the summation of a (weighted) *pure-tone dissonance* across all pairs of *tonal components* that comprise a complex sound, i.e., all (harmonic and non-harmonic) partials of all simultaneous tones combined.

Formally, given a set $\mathcal{P} = \{(f_1, a_1), ..., (f_K, a_K)\}$ of K pairs of tonal components with frequency f in Hz and amplitude a, SD can be calculated with

$$D(\mathcal{P}) := \sum_{\substack{(f_i,a_i)\in\mathcal{P} \\ (f_j,a_j)\in\mathcal{P}}} w(a_i, a_j)\, d(f_i, f_j), \tag{1}$$

[1] https://audiolabs-erlangen.de/resources/MIR/2025-ISMIR-SD

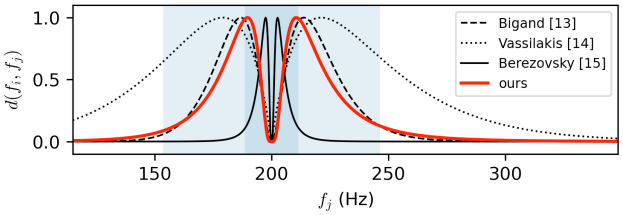

Figure 2. Dissonance kernels $d(f_i, f_j)$ for $f_i = 200$ Hz. Blue shaded areas indicate 0.25 and 1 ERB around f_i.

where $w : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ is an amplitude-dependent weighting factor and $d : \mathbb{R}_+ \times \mathbb{R}_+ \to [0, 1]$ is a model for the perceived dissonance between two pure tones. In the following, we provide an intuitive explanation of these two terms. Further details can be found on our supplemental website [1] and in [9].

The *dissonance kernel* d is based on perceptual experiments with sinusoids [11] and it should attain a high value when the frequency distance is small but not too small. Fig. 2 shows possible realizations of dissonance kernels [13–15] around $f_i = 200$ Hz. While these kernels share a similar shape, they differ in the position of the dissonance maximum. For example, for the frequency range considered in Fig. 2, the kernel by Vassilakis et al. [14] (dotted line) leads to much wider intervals with high dissonance compared to the narrow kernel used by Berezovsky [15] (solid line). The experiments in [11] suggest a maximum of perceived dissonance at around 0.25 *critical bands*, a measure for the resolution of auditory perception, often approximated in terms of *equivalent rectangular bandwidth* (ERB, e.g., [18]). Therefore, we explicitly set the maximum to 0.25 ERB, using the mean frequency $(f_i + f_j)/2$ of the pairing to determine this bandwidth. The resulting kernel is shown as a red curve in Fig. 2 and also visualized with dotted lines around each harmonic in Fig. 1c. Here, the dependency of the kernel shape on frequency is also visible, with wider kernels for higher harmonics on the linear frequency axis.

The weighting factor w ensures that pairings with high amplitudes contribute more to $D(\mathcal{P})$. We use the minimum of the two amplitudes as in [15], which is proportional to the amplitude fluctuation of the beating that occurs between the two sinusoids [9]. Additionally, many models include an exponent < 1 to account for the nonlinearity of loudness perception [8, 9, 15]. However, since the behavior of exponential compression changes when all amplitudes are scaled by a constant factor, we use logarithmic compression, resulting in the weighting factor $w(a_i, a_j) = \log\big(1 + \min(a_i, a_j)\big)$, where adding 1 ensures positive values for w [19]. Finally, it is common to also include some kind of normalization that makes $D(\mathcal{P})$ independent of the overall loudness of the sound. We omit this step here and consider loudness normalization an optional preprocessing step in Section 2.2.

While Eq. 1 allows for a wide range of possible configurations, the conceptual approach to measuring SD in music recordings remains the same. In the following, we focus on the specific parametrization described above as one illustrative example.

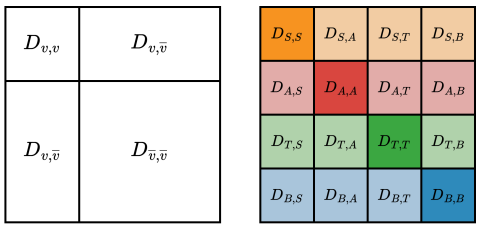

Figure 3. Illustration of considered subsets for relative SD.

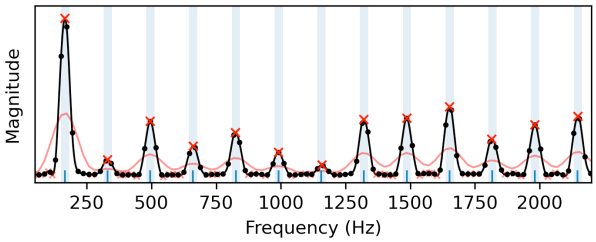

Figure 4. Illustration of harmonic peak picking for a single frame of a bass clarinet signal, showing the original DFT spectrum (black dots), the cubic spline interpolation (black line), extremal points of the interpolation function (red crosses), search range (light blue areas) around F0 multiples (blue ticks), and the local threshold (red line).

2.1 Relative Sensory Dissonance

Extending ideas from [9] and [17], we can decompose $D(\mathcal{P})$ into contributions from different components of a sound. In a multi-track polyphonic music scenario, this may give additional insights into how individual sources or voices contribute to dissonance. To formalize the decomposition, we define subsets of $\mathcal{P}$ corresponding to different voices in a sound, so that for an index set $\mathcal{V}$,

$$\mathcal{P} = \bigcup_{v \in \mathcal{V}} \mathcal{P}_v. \tag{2}$$

Furthermore, for $v \in \mathcal{V}$, we define the complementary set

$$\mathcal{P}_{\overline{v}} = \bigcup_{v' \in \mathcal{V} \setminus \{v\}} \mathcal{P}_{v'} \tag{3}$$

containing the tonal components of all other voices. Extending Eq. 1, we compute the SD between the tonal components from two separate sets $\mathcal{P}_1$ and $\mathcal{P}_2$ with

$$D(\mathcal{P}_1, \mathcal{P}_2) := \sum_{\substack{(f_i, a_i) \in \mathcal{P}_1 \\ (f_j, a_j) \in \mathcal{P}_2}} w(a_i, a_j)\, d(f_i, f_j), \tag{4}$$

which we also denote by $D_{1,2}$ in the following for brevity, along with D without subscripts for the overall dissonance $D(\mathcal{P}, \mathcal{P})$. Eq. 4 allows us to decompose D for any voice $v \in \mathcal{V}$ as shown in Fig. 3 on the left. Here, $D_{v,v}$ represents the *intrinsic* SD within a single voice, $D_{\overline{v},\overline{v}}$ is the SD independent of v, and $D_{v,\overline{v}}$ is the *relative* SD between a voice and its accompaniment. This results in a final decomposition $D = D_{v,v} + 2D_{v,\overline{v}} + D_{\overline{v},\overline{v}}$.

As a concrete example, consider the four-part chorale case, where $\mathcal{V} = \{S, A, T, B\}$. The right side of Fig. 3 illustrates the possible relative SD pairings in this scenario. Particularly relevant are the intrinsic dissonances $D_{S,S}$, $D_{A,A}$, $D_{T,T}$, and $D_{B,B}$, as well as the relative dissonances $D_{S,\overline{S}}$ (depicted in orange), $D_{A,\overline{A}}$ (red), $D_{T,\overline{T}}$ (green), and $D_{B,\overline{B}}$ (blue). It becomes evident from the right side of Fig. 3 that the individual $D_{v,\overline{v}}$ are not independent, since a change in any $\mathcal{P}_v$ influences the relative SD measurements for all other voices due to the symmetry of the matrix. Yet, Fig. 1d shows how $D_{v,\overline{v}}$ can still be used to clearly identify the different contributions to overall SD.

2.2 Tonal Components in Recordings

So far, we have only considered the case where $\mathcal{P}$ represents a sound with constant characteristics. To account for time-varying music signals, let $\mathcal{P}(m)$ represent the tonal components (equivalent to harmonics in the monophonic case) at a time index $m \in \mathbb{Z}$. In a multi-track scenario, we further require a method to obtain robust estimates of $\mathcal{P}_v(m)$ from a time-domain signal x_v for $v \in \mathcal{V}$. The frequency resolution of a discrete Fourier transform (DFT) with any reasonable window size would not be sufficient for this purpose. Prior methods have used a peak picking algorithm on the magnitude spectrum, refined with parabolic interpolation [20] or by zero-padding [9], which requires intricate fine-tuning of parameters for robust detection of harmonics. Instead, we propose a targeted peak picking approach that leverages fundamental frequency (f_0) estimates, assuming a quasi-harmonic overtone structures. f_0 estimates can be obtained robustly using monophonic algorithms (e.g., [21]), or with predominant [22] or polyphonic [23] f_0 estimation algorithms for more complex input signals. Since many musical instruments produce near-harmonic spectra, this method is applicable to a wide variety of recordings.

The frequency and amplitude of each harmonic is obtained in three steps, as illustrated in Fig. 4. First, we construct a cubic interpolating spline representation (black curve) of the magnitude spectrum (black dots) and compute its derivative, a piecewise quadratic function. Second, we find the roots of the spline derivative, identifying the extremal points of the original cubic splines (red crosses). The largest extremum in the vicinity of nf_0 (blue rectangles) is then assumed to be the spectral peak corresponding to the nth harmonic. To refine these estimates (e.g., when the even harmonics of a clarinet are below background noise level), we finally apply local thresholding, removing any peaks whose total magnitude is below a certain value. The local threshold (red curve) is calculated using a sliding Hann window with an adaptive width depending on f_0, so that the averaging spans $N_w = \lceil 1.2 \cdot f_0 N / f_\mathrm{s} \rceil$ frequency bins. This way, a single harmonic with large magnitude does not influence the threshold at the neighboring harmonics. With this method, it is possible to accurately track harmonics across multiple frames of an STFT without explicitly modeling continuity between frames. For STFT frames where the f_0 estimate indicates an unvoiced frame, we set $\mathcal{P}_v(m) = \emptyset$.

When using logarithmic compression for w in Eq. 1, the scale of the amplitudes a_i in $\mathcal{P}_v(m)$ can be arbitrary,

as long as it is consistent across all voices. It can further be desirable to remove the influence of loudness (and also relative level differences between the individual voices), which can be achieved by normalizing all amplitudes in $\mathcal{P}_v(m)$ to sum to one. In the particular case of ChoraleBricks voices are recorded in isolation and not in a musically meaningful loudness balance anyway. Therefore, we divide each a_i by a constant

$$C = \max\big(c, \sum_{(f_i,a_i)\in\mathcal{P}_v} a_i\big), \tag{5}$$

where limiting C to not become smaller than $c = 0.1$ avoids artificially inflating the influence of frames with very low overall level.

3. SENSORY DISSONANCE IN CHORALEBRICKS

A main goal of this paper is to study the behavior of SD in a realistic musical scenario. Specifically, we want to disentangle the influence of different musical properties—timbre, tuning, and score—of individual voices in a multitrack recording on the SD values, and to which extent this measure enables comparisons between different instruments, takes, or compositions.

For this exploratory analysis, we use the ChoraleBricks dataset [2], which provides recordings of ten Baroque four-part chorales. Their composition style is *homophonic*, i.e., voices follow a synchronized rhythm while forming chords with a main melody (usually in the soprano), using relatively simple chords and voicings. Furthermore, the dataset contains isolated recordings of each voice played on several different wind instruments, allowing for a comparison of various four-instrument combinations (*ensembles*) playing the same chorale. To simplify the notation of ensembles, we introduce shorthands. The ensemble used in most experiments and visualizations (e.g., Fig. 1) is denoted by $E = (\texttt{tp}, \texttt{cl}, \texttt{bar}, \texttt{bs})$, using a tuple of instrument IDs as shown in Table 1 in the order of S, A, T, and B. To denote a single instrument being replaced in E, we use a subscript, for example, $E_{\mathrm{S}=\texttt{fl}} = (\texttt{fl}, \texttt{cl}, \texttt{bar}, \texttt{bs})$. In addition to E (with two woodwinds and two brass instruments), we assemble a pure woodwinds ensemble $E_{\mathrm{wood}} = (\texttt{ob}, \texttt{cl}, \texttt{bs}, \texttt{bcl})$ and a pure brass ensemble $E_{\mathrm{brass}} = (\texttt{tp}, \texttt{fh}, \texttt{bar}, \texttt{bar})$, always choosing the instrument with the highest number of available recordings for the respective voice. Finally, as a synthetic baseline for comparison, we create an ensemble E_{saw}, where each voice is synthesized using a sawtooth waveform with 20 harmonics and the respective 12-tone equal temperament (12-TET) frequency for each note as F0. Using these ensembles and the excerpt from Fig. 1 as a running example, we can try to disentangle the influences of timbre, tuning and score on SD.

3.1 The Influence of Timbre

In Fig. 1e, large differences between instruments in relative (lighter color) and intrinsic SD (darker color) can be observed, most prominently for the T and B voice, played on `bar` and `bs`. To reveal the cause of these differences, we first characterize the timbre of the different instruments. While the full phenomenon of timbre encompasses many properties of a sound [24], we focus on the the most relevant aspect for SD, namely the distribution of amplitude across harmonics independent of the overall loudness of the sound, as expressed by the normalized $\mathcal{P}_v$. Fig. 5 shows the average amplitude distribution in the first 20 harmonics for each instrument in ChoraleBricks. Woodwinds (except for the flute) tend to have higher values in the upper harmonics (a *brighter* timbre), while flute, fluegelhorn and baritone horn on average have most sound energy concentrated in the low harmonics. We can quantify this difference with the *harmonic centroid* (HC), describing the mean harmonic index weighted by average amplitude, similar to the spectral centroid, which is often used as a timbre descriptor [25]. The HC for each instrument is given in Table 1, and we can observe that for example the bass clarinet (`bcl`) has its energy centered around the seventh harmonic (HC of 6.98), while for the flute (`fl`), the HC indicates that the most energy is in the fundamental (HC of 1.55).

These differences in timbre naturally also affect both

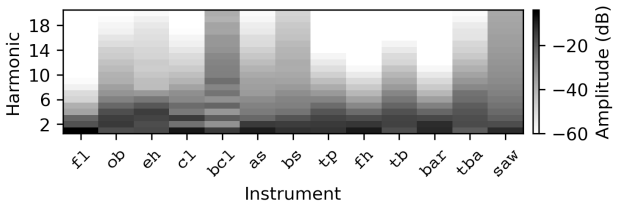

Figure 5. Amplitude distribution across the first 20 harmonics in $\mathcal{P}_v$ for each instrument (averaged over the entire datset, not showing values below -60 dB). The sawtooth timbre `saw` is shown for comparison.

Instrument	ID	# Notes	$D_{v,v}$	$D_{v,\overline{v}}$	HC
Flute	`fl`	449	0.03	1.85	1.55 ± 0.3
Oboe	`ob`	449	0.60	3.65	3.67 ± 0.7
English Horn	`eh`	460	0.74	3.70	3.78 ± 0.8
Clarinet	`cl`	909	0.53	3.43	2.94 ± 0.7
Bass Clarinet	`bcl`	545	4.60	6.08	6.98 ± 2.0
Alto Sax	`as`	103	0.94	4.38	3.06 ± 0.9
Baritone Sax	`bs`	816	1.92	4.68	3.99 ± 1.4
All Woodwinds		3731	1.34	3.97	3.71
Trumpet	`tp`	909	0.22	3.35	3.00 ± 0.8
Fluegelhorn	`fh`	909	0.12	2.75	2.25 ± 0.6
Trombone	`tb`	372	0.45	2.83	3.69 ± 1.3
Baritone	`bar`	963	0.16	1.90	2.43 ± 0.8
Tuba	`tba`	464	1.12	2.59	4.50 ± 1.3
All Brass		3668	0.41	2.68	3.17
Sawtooth S		449	1.37	5.76	5.56 ± 0.0
Sawtooth A		460	1.43	7.01	5.56 ± 0.0
Sawtooth T		456	1.50	7.02	5.56 ± 0.0
Sawtooth B		464	1.71	4.92	5.56 ± 0.0
All Sawtooth		1829	1.50	6.18	5.56

Table 1. SD statistics by instrument, showing intrinsic SD $D_{v,v}$, relative SD $D_{v,\overline{v}}$, and the harmonic centroid (HC, mean $\pm$ standard deviation). Values for the synthetic sawtooth ensemble E_{saw} shown for comparison.

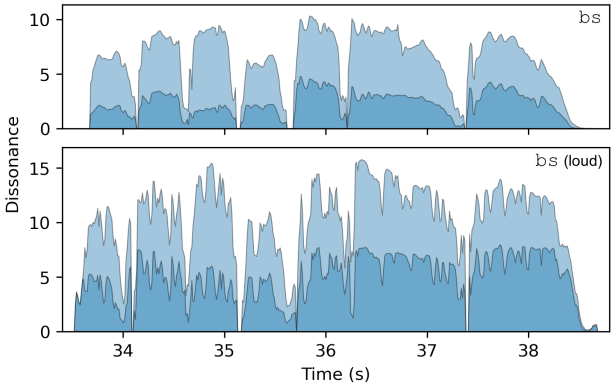

Figure 6. Two different takes for B played with `bs` (performed normally and loudly) for the excerpt from Fig. 1. Note that both $\mathcal{P}_v$ are loudness-normalized, so that differences in D only stem from changes in timbre and tuning.

intrinsic SD $D_{v,v}$ and relative SD $D_{v,\overline{v}}$. Table 1 also lists the average $D_{v,v}$ and $D_{v,\overline{v}}$ for each instrument, which are calculated relative to E, i.e., by replacing the respective voice in E with this specific instrument. For example, when the trombone played both T and B, we consider $E_{T=\texttt{tb}}$ and $E_{B=\texttt{tb}}$ for each chorale where the respective recording is available. We can observe that HC and the SD values are highly correlated. Furthermore, the instrument choice accounts for 49% of the variation in $D_{v,\overline{v}}$ over the entire dataset and 78% of the variation in $D_{v,v}$, as indicated by the effect size (η^2) for the Kruskal-Wallis nonparametric statistical test [26]. This makes the instrument's harmonic amplitude distribution the largest predictor of a voice's contribution to SD.

In addition, there are also considerable timbre variations *within* each instrument, reflected for example in the standard deviation of HC reported in Table 1. As an illustrative example, Fig. 6 shows two different takes of the B voice played on `bs`. In the second take ("loud"), the player was instructed to play as loud as possible. Notably, despite the loudness normalization of $\mathcal{P}_v$, both $D_{v,v}$ and $D_{v,\overline{v}}$ are larger by almost a factor of two for the loud take, which also shows stronger fluctuations of relative SD within the individual chords. The possibility to influence SD through variations in an instrument's timbral qualities may also explain the reduction in SD towards the end of the example in Fig. 1, an effect that is not present for the loud take.

3.2 The Influence of Tuning

SD has previously been shown to be a context-sensitive measure for tuning related to just intonation (JI) [16, 17]. In this section, we aim to quantify how large the effect of tuning is on SD compared to other influences like timbre. To visualize this in our running example, we *virtually pitch-shift* each voice by a certain amount $p \in [-50, 50]$ in cents by multiplying all frequencies in $\mathcal{P}_v(m)$ with a factor $2^{p/1200}$. We then calculate for each value of p the new relative dissonance $D_{v,\overline{v}}$ against the unmodified other voices of the ensemble ($D_{v,v}$ remains mostly unaffected by a small pitch shift). In other words, we measure how much relative SD changes when a performer would change their

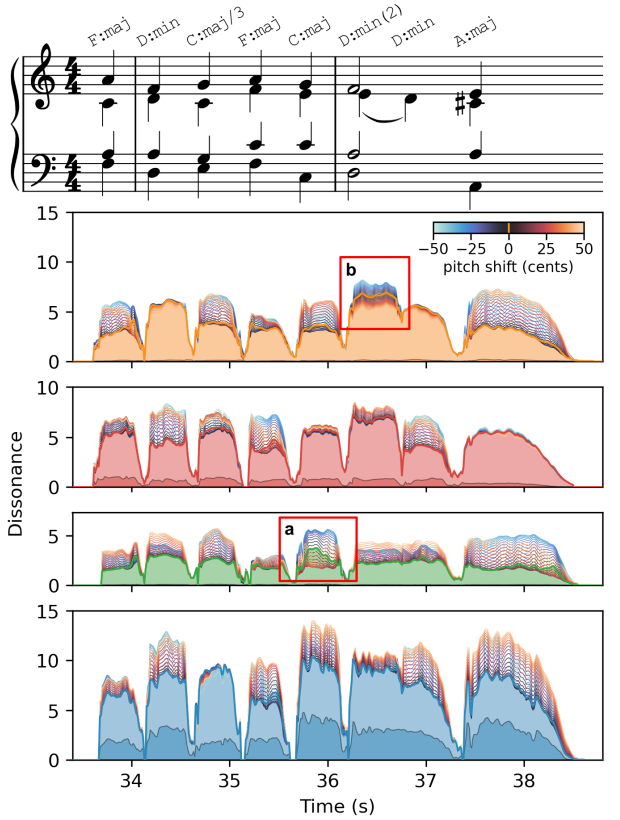

Figure 7. Influence of detuning each voice in E by ± 50 cents (excerpt from `GE1`, as in Fig. 1). Chord labels and sheet music are shown time-aligned for reference.

intonation while the rest of the ensemble plays normally.

The result is shown in Fig. 7. $D_{v,v}$ and $D_{v,\overline{v}}$ is plotted for each voice separately, and the lines from light blue ($p = -50$ cents) to light red ($p = 50$ cents) indicate the change in $D_{v,\overline{v}}$ with variable tuning. Notably, the *tuning sensitivity* of $D_{v,\overline{v}}$ (i.e, the range between the maximum and minimum $D_{v,\overline{v}}$ within a note when p is varied by ± 50 cents) changes between notes and chords. In particular, for most major and minor thirds in the excerpt, SD increases only slightly even for the largest pitch shifts. Considering the statistics over the entire dataset, we find that for the ensemble E, the root note has an average tuning sensitivity of 2.71, the minor third of 1.09, the major third of 1.01, and the fifth of 1.70 (only considering chord degrees with frequent occurrences). We must however account for the fact that the root note of each chord is often doubled (in the same or a different octave) in the four-part voicings of chorales. If we consider only cases where the respective note is not doubled, we record an average tuning sensitivity of 1.13 (root), 0.75 (minor third), 0.87 (major third), and 1.62 (fifth). This suggests that SD is more sensitive to the tuning of the root and fifth than to that of the thirds. Furthermore, in comparison to the mean relative SD by instrument in Table 1, tuning sensitivity in general is relatively small, so that measuring tuning (e.g., 12-TET and JI, which often differ only by a few cents) with SD is only meaningful in conditions where timbre remains unchanged.

There are two cases in the excerpt where the visualization shows that a change in tuning would significantly reduce relative SD. First, the root note of the `C:maj` chord

Chord Type	#	D			
		E	E_{wood}	E_{brass}	E_{saw}
maj	229	15.21	31.74	7.90	29.73
min	119	15.99	34.14	9.02	31.96
maj/3	55	17.65	36.16	10.50	31.65
sus4	16	16.64	34.75	8.72	30.02
min/3	12	19.17	36.52	7.83	27.78
others	32	17.00	33.11	10.25	31.20

Table 2. SD statistics by different chord types. Chord symbols follow the notation scheme from [28].

in the T voice (red box **a** in Fig. 7) would contribute less to D if it was played around 25 cents higher. In fact, according to the annotated F0, the T voice is on average 17 cents flat relative to the octave formed with B. Interestingly, this does not affect the "optimal tuning" (w.r.t. SD) of B, indicating that the intervals formed with S and A are stable. Second, for the minor third of the `D:min(2)` chord in S (red box **b** in Fig. 7), SD would be reduced if the note was played up to 50 cents lower. Given that this note forms the musically dissonant interval of a minor second with A, this is an indication for a case where SD does not exhibit a local minimum near the interval prescribed by the score. This can become a problem for approaches where SD is used as a measure to "optimize" tuning [16, 17].

3.3 The Influence of Score

Finally, we consider the influence of score on SD, in particular by analyzing differences between chords. In Fig. 7, a prominent example for the score influence is the `D:min(2)` chord that has the highest overall SD in the excerpt. The relative SD of S, A, and B, as well as the overall SD significantly drops when the musically dissonant suspended second is resolved in voice A.

Since the chord vocabulary of the compositions is relatively limited (major and minor chords in root position account for 75% of all chords), we can statistically analyze differences between chord types by grouping them as shown in Table 2, using the chord labels without root. In particular, we compare how the chord type influences the overall SD for E, E_{wood}, E_{brass}, and E_{saw} by computing the mean D over the entire dataset. In fact, the variations in mean SD by chord type are mostly consistent across ensembles. For example, the mean SD of `maj` chords is lower than that of `min` chords. /3 chords with the third in B are more dissonant, except for `min/3` in E_{brass}. These trends partly resemble a ranking based on subjective dissonance ratings of triads and tetrads [27], replicating a previous result with synthetic data [7]. However, even within one ensemble, we find that the chord type only accounts for between 6% (E_{wood}) and 11% (E_{brass}) of the variation in D (according to the Kruskal-Wallis test as above).

Another score-related property of each chord is its voicing, i.e., the assignment of notes to the individual voices. Two effects can be observed in the statistics for individual voices of E_{saw} in Table 1. First, the average $D_{v,v}$ increases towards lower voices. This follows from more harmonics falling into the critical bands around neighboring harmonics, as can be observed in Fig. 1b. Second, the average $D_{v,\overline{v}}$ is higher for middle voices, indicating that these voices tend to contribute more to the overall SD, independent of timbre and tuning, which is fixed in E_{saw}.

4. DISCUSSION & APPLICATION EXAMPLES

From this analysis, we can draw three main conclusions for using SD as an informative measure in music production. First, the relative SD of individual tracks provides insights into their musical interaction, e.g., in terms of tuning and the musical dissonance of intervals. However, as an absolute measure, SD is mainly determined by instrument timbre, prohibiting direct comparisons across instrument classes. Second, SD may offer an advantage over F0-based tuning analysis because it is context-sensitive, accounting for intervals between all voices. The sensitivity to tuning differences depends on the chord degree and voicing, where even strongly detuning a third by ± 50 cents may in some chords only slightly affect the SD measure. Third, while systematic score influences are present, they are comparatively small, which makes applications like musicological analysis as outlined in [9] only feasible within controlled scenarios. This suggests a number of practical applications for SD measures, two of which we briefly want to outline in the following.

Take Selection: Comparing the relative SD between multiple takes of the same excerpt played with the same instrument, like in Fig. 6, can indicate differences in terms of timbre and/or tuning in relation to a specific accompaniment. Together with other musically motivated quality measures (e.g., [29]), this could serve as the basis for suggesting an edit sequence that aligns best with the desired properties of the track. As an example, the fifth note of the T voice in the excerpt from Fig. 1 could be replaced with a version from a different take that is more in tune compared to the other voices. Since in classical music production, sound engineers sometimes have to deal with dozens of takes for a single passage, even sorting the takes by SD in a certain chord could aid the selection process.

Equalization: Since timbre is the primary contributing factor to relative SD, we can aim to modify it through equalization to decrease (or increase) SD while still preserving the instrument's characteristic sound. As an example, applying a filter to the B voice in the excerpt from Fig. 1, reducing magnitudes by only 6 dB between 2 and 4 kHz, leads to a reduction of the mean $D_{v,\overline{v}}$ by 11% (from 3.53 to 3.15) within the excerpt. This adaptation could also be made context-dependent, e.g., by reducing the amplitude of harmonics that contribute strongly to SD only when interfering voices are present in the mix.

Finally, it should be emphasized again that SD is not measuring the actual perceived dissonance and that variations in the model assumptions for SD, as well as other acoustic measures like those based on harmonicity (e.g., [6, 30]), could yield significantly different, but equally valid results. Understanding the properties of different dissonance measures in realistic musical scenarios will be key to establishing them as a music production tool.

5. ACKNOWLEDGEMENTS

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Grant No. 401198673 (MU 2686/13-2) and 555525569 (MU 2686/18-1). The International Audio Laboratories Erlangen are a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and Fraunhofer Institute for Integrated Circuits IIS.

6. REFERENCES

[1] J. Tenney, *A History of 'Consonance' and 'Dissonance'*. New York: Excelsior Music Publishing Company, 1988.

[2] S. Balke, A. Berndt, and M. Müller, "ChoraleBricks: A modular multitrack dataset for wind music research," *Transaction of the International Society for Music Information Retrieval (TISMIR)*, vol. 8, no. 1, pp. 39–54, 2025.

[3] T. Eerola and I. Lahdelma, "The anatomy of consonance/dissonance: Evaluating acoustic and cultural predictors across multiple datasets with chords," *Music & Science*, vol. 4, pp. 1–19, 2021.

[4] J. H. McDermott, A. F. Schultz, E. A. Undurraga, and R. A. Godoy, "Indifference to dissonance in native amazonians reveals cultural variation in music perception," *Nature*, vol. 535, no. 7613, pp. 547–550, 2016.

[5] W. A. Sethares, "Local consonance and the relationship between timbre and scale," *Journal of the Acoustical Society of America*, vol. 94, no. 3, pp. 1218–1228, 1993.

[6] F. Stolzenburg, "Harmony perception by periodicity detection," *Journal of Mathematics and Music*, vol. 9, no. 3, pp. 215–238, 2015.

[7] P. M. C. Harrison and M. T. Pearce, "Simultaneous consonance in music perception and composition," *Psychological Review*, vol. 127, no. 2, pp. 216–244, March 2020.

[8] R. Marjieh, P. M. C. Harrison, H. Lee, F. Deligiannaki, and N. Jacoby, "Timbral effects on consonance disentangle psychoacoustic mechanisms and suggest perceptual origins for musical scales," *Nature Communications*, vol. 15, p. 1482, 2024.

[9] W. A. Sethares, *Tuning, Timbre, Spectrum, Scale*. London: Springer, 1998.

[10] H. L. F. Helmholtz, *On the Sensations of Tone as a Physiological Basis for the Theory of Music*, 4th ed. Longmans, Green, and Co., 1912.

[11] R. Plomp and W. J. M. Levelt, "Tonal consonance and critical bandwidth," *Journal of the Acoustical Society of America*, vol. 38, no. 4, pp. 548–560, 1965.

[12] W. Hutchinson and L. Knopoff, "The acoustic component of western consonance," *Interface*, vol. 7, no. 1, pp. 1–29, 1978.

[13] E. Bigand, R. Parncutt, and F. Lerdahl, "Perception of musical tension in short chord sequences: The influence of harmonic function, sensory dissonance, horizontal motion, and musical training," *Perception & Psychophysics*, vol. 58, no. 1, pp. 125–141, 1996.

[14] P. N. Vassilakis and R. A. Kendall, "Psychoacoustic and cognitive aspects of auditory roughness: definitions, models, and applications," in *Proceedings of Human Vision and Electronic Imaging XV*, vol. 7527. Bellingham, WA: SPIE, 2010.

[15] J. Berezovsky, "The structure of musical harmony as an ordered phase of sound: A statistical mechanics approach to music theory," *Science Advances*, vol. 5, no. 5, p. eaav8490, 2019.

[16] W. Sethares, "Adaptive tunings for musical scales," *The Journal of the Acoustical Society of America*, vol. 96, 1994.

[17] S. Schwär, S. Rosenzweig, and M. Müller, "A differentiable cost measure for intonation processing in polyphonic music," in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Online, 2021, pp. 626–633.

[18] B. R. Glasberg and B. C. J. Moore, "Derivation of auditory filter shapes from notched-noise data," *Hearing Research*, vol. 47, no. 1-2, pp. 103–138, 1990.

[19] M. Müller, *Fundamentals of Music Processing – Using Python and Jupyter Notebooks*, 2nd ed. Springer Verlag, 2021.

[20] J. O. Smith and X. Serra, "PARSHL: An analysis/synthesis program for non-harmonic sounds based on a sinusoidal representation," in *Proceedings of the International Computer Music Conference (ICMC)*. Computer Music Association, 1987.

[21] A. Camacho and J. G. Harris, "A sawtooth waveform inspired pitch estimator for speech and music," *The Journal of the Acoustical Society of America*, vol. 124, no. 3, pp. 1638–1652, 2008.

[22] J. Salamon, E. Gómez, D. P. W. Ellis, and G. Richard, "Melody extraction from polyphonic music signals: Approaches, applications, and challenges," *IEEE Signal Processing Magazine*, vol. 31, no. 2, pp. 118–134, 2014.

[23] C. Weiß and G. Peeters, "Comparing deep models and evaluation strategies for multi-pitch estimation in music recordings," *IEEE/ACM Transactions on Audio, Speech & Language Processing*, vol. 30, pp. 2814–2827, 2022.

[24] K. Siedenburg, C. Saitis, and S. McAdams, Eds., *Timbre: Acoustics, Perception, and Cognition*, ser. Springer Handbook of Auditory Research. Cham, Switzerland: Springer, 2019, vol. 69.

[25] G. Peeters, "A large set of audio features for sound description (similarity and classification) in the CUIDADO project," IRCAM, Paris, France, Tech. Rep., 2004.

[26] W. H. Kruskal and W. A. Wallis, "Use of ranks in one-criterion variance analysis," *Journal of the American Statistical Association*, vol. 47, no. 260, pp. 583–621, 1952.

[27] D. L. Bowling, D. Purves, and K. Z. Gill, "Vocal similarity predicts the relative attraction of musical chords," *Proceedings of the National Academy of Sciences*, vol. 115, no. 1, pp. 216–221, 2018.

[28] C. Harte, M. B. Sandler, S. Abdallah, and E. Gómez, "Symbolic representation of musical chords: A proposed syntax for text annotations," in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, London, UK, 2005, pp. 66–71.

[29] O. Romani Picas, H. Parra Rodriguez, D. Dabiri, H. Tokuda, W. Hariya, K. Oishi, and X. Serra, "A real-time system for measuring sound goodness in instrumental sounds," in *Proceedings of the 138th Audio Engineering Society Convention (AES)*, Warsaw, Poland, May 2015, pp. 1106–1111.

[30] P. M. C. Harrison and M. T. Pearce, "An energy-based generative sequence model for testing sensory theories of western harmony," in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 2018, pp. 160–167.