Quantifying sound colour of musical instruments – precise harmonic timbre coordinates of like instruments

Rok Prislan\textsuperscript{1,2}, Urša Kružič\textsuperscript{3}, and Daniel Svenšek\textsuperscript{4,5,*}

\textsuperscript{1} Acoustic Laboratory, InnoRenew CoE, Livade 6a, 6310 Izola, Slovenia
\textsuperscript{2} University of Primorska, Faculty of Mathematics, Natural Sciences and Information Technologies, Glagoljaška 8, 6000 Koper, Slovenia
\textsuperscript{3} Music School Vrhnika, Trg Karla Grabeljška 3, 1360 Vrhnika, Slovenia
\textsuperscript{4} Department of Physics, Faculty of Mathematics and Physics, University of Ljubljana, Jadranska 19, 1000 Ljubljana, Slovenia
\textsuperscript{5} Laboratory of Molecular Modeling, National Institute of Chemistry, Hajdrihova 19, 1000 Ljubljana, Slovenia

Received 7 May 2023, Accepted 24 December 2023

Abstract – Timbre – sound “colour” – is an abstract, delicate property of sound, especially in a high-value context such as musical instruments. It is a perceptual construct so intangible that it cannot be considered a quantity. Since sound nevertheless reaches our ears as a complete physical reality, we hypothesize that this inherent abstraction of its timbre is primarily due to the lack of a meaningful, musically relevant, and robust quantification that would do justice to the subtlety of human auditory perception. It is therefore not surprising that not a single aspect of timbre is to be found in the specifications of musical instruments. We introduce harmonic timbre coordinates, concrete and robust numbers that quantify a partial aspect of timbre of an instrument’s sound – its harmonic structure – with a precision that allows relevance in the musical context. These numbers could, for example, help a buyer find an instrument whose sound is closer to his or her preferences. Or they could enable precise tracking of harmonic changes in sound, and more.

Keywords: Musical instruments, Harmonic timbre coordinates, Robust and precise measure of harmonic timbre, Linear harmonic timbre vector space

1 Introduction

“Maybe you should try the instrument I got to hear the other day – I would definitely describe it as Italian sounding, perhaps more along the lines of Guarneri than Stradivari.”

“Very interesting indeed, but I tend to prefer a slightly bolder German sound, especially with a touch of French warmth in it…” Have you ever heard a similar conversation between musicians? This one is from the world of strings and is, of course, an indecent cliché. Nevertheless, the abstract psychoacoustic aspects of sound colour – timbre [1] – are generally qualified and conveyed in this way – and in countless other, less “national” ways. It goes without saying that the designations presented here need have nothing to do with the actual country of origin of an instrument. They are almost archetypal descriptors that have evolved over the centuries of instrument making practice, probably based on best examples taken as ideals. Other descriptors of timbre range from subtle, highly subjective impressions on one hand to more tangible descriptions like brightness, thickness, nasality, etc. [2, 3] on the other. There are also rigorously defined “elementary auditory sensations” such as roughness, sharpness, tonality, and loudness [4, 5] that would likely be widely agreed upon in listening tests but are too prosaic to describe delicate timbre differences between like instruments. They are useful for distinguishing between families of instruments (strings, woodwinds, brass, etc.), and between different members of a family, e.g., between a clarinet and an oboe [6–10]. However, this is not what we have in mind. One wants to help the two musician friends above, which is much more ambitious. And countless buyers of instruments, professional and amateur musicians, aspiring young musicians, their parents, and so on.

To be fair and clear, a scientific method is always partial, especially in areas like the present one that touches on humanistic questions. It can hardly ever replace actual listening to the instrument, let alone trying it out (because the playability\textsuperscript{1} [2, 3] of an instrument is another quality besides its sound). But it can serve as a starting point,
a first guide according to one’s preference, to conveniently narrow down the rich and dazzling selection. Imagine the impact that labelling instruments with relevant timbre quantifiers – actual numbers – could have on the production and market of musical instruments. The harmonic timbre coordinates we are about to introduce are potentially such numbers, describing harmonic sound colour as a small aspect of the holistic musical sense of timbre [1].

If these coordinates are to have real value in the musical sense and not be merely a scientific or technical curiosity, they must be precise enough to describe the minute differences between like instruments, i.e. different instruments of the same type, and extremely robust to be reproducibly determined from these minute differences. In other words, for the harmonic aspects of timbre addressed, the method must reach the limits of musicians’ perception. This is a challenge. The goal of this work is a proof of concept – to test whether robust quantification of sound colour aspects is possible, including all steps: playing, recording, and analyzing. A particular challenge is that the instruments should be played by a musician (rather than artificially excited, which would be less prone to variation) and handled by the musician in the usual way, with small variations in geometric positioning in space with respect to the microphones and other naturally present influences.

Moreover, the coordinates of a given instrument should not be a fingerprint of that instrument, distinguishing it as sharply as possible from any other. Well-developed methods from the worlds of speech processing [11–14] and instrument classification [8, 9, 15–22] would lend themselves to such a forensic goal. The latter aim to automatically identify different instruments and possibly separate them in a polyphonic audio environment. With the exception of “eigeninstrument” [15–17], timbre trajectory tracking [23, 24] and formant [25] approaches, which are closer to us, they are usually based on a large set of mel-frequency cepstral coefficients (MFCC) and other computed spectro-temporal quantifiers [7, 8, 14, 19, 21, 22, 26–28], which have only an indirect and especially one-way connection to sound – they can be determined from sound, but not vice versa. In contrast, a musically relevant harmonic timbre coordinate space should have a direct two-way relation to sound. It should also have a natural metric that defines a meaningful timbral distance, so that instruments whose coordinates are close together sound similar, and instruments that are farther apart in the coordinate space sound different [25].

The studies of Refs. [29, 30] are most related to our approach, investigating harmonic overtones of violins and extracting the tendencies of instruments from the spectral envelope. It was shown [30] that listeners could detect the difference in harmonic timbre when there were differences in the tendencies of the spectral envelopes. This speaks to the use of our approach, which also quantifies the harmonic content, but with the important difference in the recording of the instrument and also the analysis part. In Ref. [30], the recordings were made with a single microphone about 10 cm above the bridge of the violin. While this results in a reproducible recording of the instrument, it does not represent a musically relevant listening position and is biased toward specific details of the acoustic nearfield that would change significantly with a slight change in the recording point. The analysis extracted the differences in spectral tendencies, while in our case a more elaborate analysis method statistically identifies the most discriminating harmonic features without prior assumptions.

2 Strategy

Our strategy is based on two fundamental principles. The first is to remain completely objective and not to include any aspects of psychoacoustics or neuroacoustics [10, 31–35], apart from the commonly known general relationship between the harmonic timbre of a sound and the structure of its power spectrum. For reasons of precision and reproducibility, we intentionally limit ourselves to strictly stationary, periodic sound. This means a perfectly harmonic spectrum consisting only of the fundamental plus overtones. This restriction could be relaxed in future developments, as far as the reproducibility of the performance and a sufficient irrelevance of the (competent) performer allow it. As it stands, the method is directly applicable to situations in which the essentially stationary part of an instrument’s tone can be sufficiently long, which is always the case for instrument families with sustained (driven) sound, such as woodwinds, brass, and strings. On the other hand, for instruments with impulsive sound, such as the piano or guitar, the method is still applicable, but probably less useful. In this case, the transients (attack, decay, release) and the temporal modulation resulting from imperfect harmonicity become more important, and if they are omitted, the description of timbre is somewhat impoverished. The present pilot study was conducted with violoncelli, which are known for their rich timbre.

The second of our fundamental principles is a top level approach, close in some ways to data mining; our input data is exclusively what is actually heard when the instrument is played as usual – its radiated sound. The instrument itself remains a black box that we do not want to analyze mechanistically. We renounce any sound or vibration mode analysis, tone generator models with delicate nonlinear phenomena like edge vortices (flute) or hair stick and slip on the bowed string. This has solid empirical reasons. Decades of dedicated, complex research on musical instruments, from rigorous and mainly conceptually relevant physics [36–42], musically relevant top-down approaches [43–45] to the problems arising in the practice of instrument making [40, 46], have proven that an instrument is simply too delicate and nonlinearly sensitive (chaotic) to allow bottom-up physical modeling with adequate, humanistic relevance. In other words, one cannot calculate how an instrument will sound, let alone select and purchase it on that basis. Moreover, many studies focusing on sound of musical instruments tend to underestimate the complexity of radiation (e.g., measurements with a single microphone, in a single point in space), which is delicate and technically demanding. However, radiation is as important as the vibration of the instrument – ultimately, it is the
radiated sound that we hear. We should be aware that the radiation patterns of musical instruments are extremely complex [47–52] and only reinforce the black box approach. A review of the violin acoustics [53] considers, among other important aspects, the vibrations and sound radiation of the body, the importance of the materials used, and perceptual studies of the instrument. The challenge of recording the sound of the violin is pointed out, for which no single microphone position could entirely be regarded as "typical", especially for far-field radiation. In fact, some sort of averaging is required, leading to the measurement becoming quite elaborate. All this, as well as the preceding and following statements in this paragraph, are fully in line with our experience, conceptual considerations and solutions.

3 Space of overtone spectra and principal harmonic timbres

The power spectra of acoustic instruments contain several tens of well-defined harmonic overtones $i$, more than 50 in the case of the lowest (C) string of the cello, Figure 1b. We consider the levels $h_i = \log_{10} P_i$ of their peak-integrated intensities $P_i$ as components of a vector $\mathbf{h}$ (harmonic vector), Figure 1c, in a multidimensional harmonic timbre vector space. In addition to a direct and practically bidirectional relation to sound, such elementary harmonic vectors have another very important property that is consistent with

2 It turns out that, contrary to most canonical textbooks and still persisting common knowledge, the perceived timbre of even a perfectly periodic tone depends not only on its power spectrum, but is influenced by the relative phases of the overtones. Although the influence is less than that of their relative intensity, it is nevertheless audible, as can be easily verified by a simple synthesis of the sound signal from its harmonic power spectrum. Since the sound radiation pattern is a complex function of both frequency and spatial direction, the relative intensities and phases of the harmonics depend on the direction of the listening point. While the solid angle average of the intensities is a well-defined quantity, the same cannot be said of the relative phases. For this reason, the relative phases of the radiated harmonics are inherently less well defined, although in the case of the vibration modality they are perfectly stable over the entire duration of the sustained tone and, moreover, over different repetitions of the same tone, i.e. they are a fingerprint of the individual (nonlinear) tone generator.
with the principles of hearing: they are by definition invariant to pitch shifts. They encode the harmonic timbre completely independently of the pitch and therefore allow a direct comparison of the timbres of tones with different pitches.

So many harmonics are far too many parameters to be reproduced stably while playing or determined reliably from playing, so that they would represent meaningful harmonic timbre coordinates. Therefore, one must identify a smaller number of relevant collective features of the overtone spectrum as a whole. These features are not known a priori and vary from one type of instrument to another. The most important features are therefore determined together with the harmonic timbre coordinates of the individual instruments on the basis of a statistically sufficiently large collection of instruments of a given type. The result is then also the relevant harmonic features - principal timbres - of that instrument type [24, 31], together with the coordinates of the individual instruments in the collection, which measure the presence (loading) of the principal timbres in each individual harmonic timbre. Once these principal timbres have been “mined”, purely from acoustic data, one can also listen to them and decide for oneself about their psychoacoustic interpretation. In addition, one can wish for an instrument in which one of the principal timbres is added with some positive weight, another perhaps with some negative weight, and search among the instruments whose coordinates reflect this. Note that a negative weight does not mean that there is something “less”, but means an opposite timbre.

The operation is accomplished by the so-called singular value decomposition (SVD) of the matrix A of all harmonic vectors \( \mathbf{h} \), which in our case are stacked as columns, \( A_{ij} = h_i^j \). These are the measured harmonic vectors of all instruments in the collection, where each instrument can and should be represented by multiple vectors corresponding to multiple repetitions of the same tone. Besides obvious improvements in averaging, the repetitions determine the statistical dispersion of the resulting harmonic timbre coordinates. Depending on the size of the collection, the matrix A has typically many more columns than rows (“wide” matrix). In standard notation, the decomposition is

\[
A = U S V^T, \quad A_{ij} = \sum_k U_{ik} S_k V_{kj}, \quad (1)
\]

where \( U \) is a square matrix of the size of the harmonic vector, \( S = \text{diag}(S_k) \) is a diagonal matrix of the same size with nonnegative values \( S_k \) ordered from largest to smallest, and \( V^T \) is a wide matrix of the same height and with width equal to the number of harmonic vectors in the collection.

An intuitive interpretation of the decomposition equation (1) is the key to understanding the actual meaning of the harmonic timbre coordinates. Consider the matrix A as a trivial transformation from the space of harmonic vector samples of the instruments (where the first sample in the collection is simply represented by the first basis vector \( \mathbf{e}_1^T = (1, 0, 0, \ldots) \), the second by \( \mathbf{e}_2^T = (0, 1, 0, \ldots) \), and so on) to the space of harmonics. Its \( j \)-th column, \( \mathbf{h}_j \), is the image of the basis vector \( \mathbf{e}_j \), i.e., the harmonic vector of the \( j \)-th sample, \( A \mathbf{e}_j = \mathbf{h}_j \). So far, so little. Now, here is the trick. The decomposition equation (1) inserts a new space between the sample and the harmonic space, called the feature or factor space – in our case, the space of the principal timbres. The product

\[
A \mathbf{e}_j = U S V^T \mathbf{e}_j
\]

can now be expressed via the intermediate basis – via the principal timbres. What is the advantage? First, we learn these timbre features from pure data without prior knowledge and without opening the black box. Second, by omitting a number of the least important features, we obtain the best possible approximation to the entire collection of \( \mathbf{h}_j \) given the remaining number of features.

The orthogonal matrix \( U \) connects the feature space and the space of harmonics: its columns are harmonic vectors of the principal timbres. The row-orthonormal matrix \( V^T \) connects the sample space and the feature space: its columns are the samples in the principal timbre basis. Thanks to the SVD, the feature space is very special indeed. The principal timbres, ordered by their importance factors \( S_k \), describe the harmonic timbre variations between the instrument samples in the most efficient way – they capture as much variation as possible with a truncated basis, i.e., with a limited number of harmonic vectors, smaller than the dimension of the harmonic space. Moreover, the variations described by any two principal timbres are uncorrelated across the entire collection of the samples, as shown by the row orthogonality of \( V^T \).

### 3.1 Harmonic timbre coordinates

When we execute the product equation (2) from right to left, \( (V^T \mathbf{e}_j)_k = V_{kj} e_k \) is the \( j \)-th sample expressed with the components \( V_{kj} \) of the projections onto the principal timbres – these are our harmonic timbre coordinates \( k \). Next, they are each multiplied by the corresponding importance factor \( S_k \). Finally, by multiplying by \( U \), we retrieve the harmonic vector \( \mathbf{h}_j \) of the \( j \)-th sample as the sum of the properly weighted harmonic vectors \( U_{ik} \) of the principal timbres \( k \), i.e. we have expanded the harmonic vector in terms of the principal timbres. Written out explicitly,

\[
h_j^k = A_{ij} = U_{i0} S_0 V_{0j}^T + U_{i1} S_1 V_{1j}^T + U_{i2} S_2 V_{2j}^T + \ldots,
\]

the numbers \( V_{0j}^T, V_{1j}^T, V_{2j}^T \) and \( V_{3j}^T \) are the first four coordinates of the \( j \)-th sample.

It is practical to enforce\(^3\) that the 0-th principal timbre vector \( U_{i0} \) has all components identical. In logarithmic space, the addition of such a vector to any harmonic vector corresponds to simple multiplication of the original

\(^3\) This can be achieved by adding a dummy vector with all components identical to the collection of harmonic vectors that enter the SVD, while ensuring that these are orthogonal to the dummy vector, i.e., normalizing them so that the geometric mean of the components of their power vectors equals 1.
(non-logarithmic) power vector by a scalar, i.e. a simple scaling of the signal amplitude. Then, since the principal timbres are orthogonal to each other, all other coordinates are automatically unaffected by amplitude scaling and represent pure harmonic timbres, while at the same time only the 0th coordinate depends on the amplitude of the signal or the choice of intensity unit. The harmonic timbre coordinates of the \( j \)th instrument sample are thus

\[
(c^j_1, c^j_2, c^j_3, \cdots) = (V^T_{1j}, V^T_{2j}, V^T_{3j}, \cdots).
\]

If the principal timbre basis of an instrument type is already established and represents a standard harmonic timbre coordinate system that should not be changed, the coordinates of a new instrument are determined by projecting its harmonic vector \( h \) onto the corresponding harmonic vectors of the principal timbres (columns of \( U \)) and dividing by the importance factors,

\[
c_k = \frac{1}{S_k} \sum_i U_{ik} h_i.
\]

4 Relevance, robustness, reproducibility

The SVD method is one of the most useful tools in data-driven science and engineering. Among countless applications in all fields, it has been used to classify timbres in acoustics [7, 8, 24, 31, 32]. It provides a necessary solid foundation for introducing and defining our harmonic timbre coordinates, but no more than that. We cannot overemphasize the critical importance of musical relevance, robustness, and reproducibility of the coordinates we will now discuss in more detail. Only when all this is achieved do harmonic timbre coordinates have their true value.

The requirements that ultimately make the concept of harmonic timbre coordinates a physical reality are therefore absolutely crucial and have evolved over several years of analytical goal-oriented experimentation.

(I) The coordinates must refer to the circumstances under which the instrument is normally used – they must be measured while the instrument is excited by a musician as she/he would normally play it. Any kind of artificial excitation, such as artificial lips, a mechanical bow or even direct mechanic/magnetic excitation of the strings, even if this would improve repeatability, is excluded.

(II) Nevertheless, the coordinates must be well-defined and should not vary too much in repeated playing of an identical tone, while all circumstances remain unchanged. This natural excitation variation defines a base tolerance, i.e., the highest possible/reasonable precision to which all other sources of errors are compared and which should ideally lie within this range. The base tolerance must be small compared to the differences between instruments. If this were not the case, the classification of instruments would be meaningless, which one would not expect based on experience. The results confirm that this presumption is clearly fulfilled.

(III) The variations due to the change of performer should be small compared to the differences between the instruments. Ideally, they should be within the base tolerance so that the performer is completely omitted as a relevant factor. If this were not the case, the coordinates would belong to the musicians and their playing, not to the instruments. Of course, the playing must be of the right quality and sufficiently similar (for string instruments, especially the cello, this concerns mainly the speed, pressure and contact point of the bow). Professional musicians can do this, and the results confirm it.

(IV) The coordinates must be robust, ideally within the base tolerance, to natural variations that occur in normal but not overly gesticulated performance. These variations are mainly due to changes in the position and orientation of the instrument during playing, which shift its sound radiation pattern. It would be unrealistic if the determined coordinates were not robust at least against minimal natural variations of this kind and the position/orientation of the instrument had to be rigidly fixed.

(V) The determination of the coordinates must be perfectly reproducible. When the complete measurement setup is reassembled and the instrument is again placed in the prescribed position and orientation, the results must remain within the base tolerance. This may seem like a trivial requirement, but in real-life acoustics things are never quite simple. Tiny nuances in tuning, changes in air temperature, humidity and CO₂ level, the humidity of the instrument’s wood, the type of bow, the use of rosin, and the tension of the bow hairs are just some of the possible influencing factors that need to be checked and possibly controlled when striving for precision.

4.1 Measurement and protocol

To meet these requirements, a recording method was developed that is consistent with standard performance and listening conditions, robust to small changes in instrument/player position, achieves high repeatability, and is largely independent of room acoustic conditions, if they are not too specific and modal effects are avoided by choosing a sufficiently large room.

Listening to a musical instrument inherently involves both direct and ambient sound. The extreme limits – anechoic and reverberation chambers – are far from realistic musical environments and should not be used. The complex radiation pattern [50–52] of the instrument’s direct sound must be averaged over an appropriately large solid angle [54]. The normally strongest first reflections from the room boundaries can be suppressed if the recording is done in a sufficiently large room and critical parts of the boundaries are covered with absorbing materials. The remaining reflections cannot be avoided and merge into a reverberant sound field, which is, however, sufficiently diffuse if the room is large enough. A diffuse field is stochastically irregular and by no means homogeneous – at any frequency it is spatially
correlated on the wavelength scale [55–57]. It therefore requires spatial averaging⁴, which is essentially statistical in nature.

The Schroeder frequency [58] of the room used for the presented recordings was 108 Hz. It is determined from the volume of the room (length 15.6 m, width 11.6 m, height 4.7 m, volume 853 m³) and the reverberation time (1.0 s at 1000 Hz) in the frequency range of interest. The room is normally used as a concert venue and has no significant sound-absorbing measures, except for a smaller number of curtains on the walls and several absorbing ceiling panels. During the measurement, highly sound-absorbing foam was added to room surfaces from which strong direct reflections emanate. While adequate for the purpose of this study, for advanced work with violoncelli larger rooms with lower Schroeder frequencies should be used when necessary to accommodate the lowest range of the cello.

Both tasks of spatial averaging of direct and ambient sound are accomplished by recording with a large number of microphones distributed over a characteristic listening solid angle in front of the instrument. After several tests with up to 60 microphones, a cross-shaped microphone stand was developed, Figure 2. With a safety margin, it includes 33 microphones⁵ evenly spaced 7.5 cm apart in a horizontal and vertical line. Because of the relatively large wavelength of the fundamental tones, the distance between the most distant microphones was set at 127.5 cm.

To avoid spatial aliasing, the spacing between the microphones (7.5 cm) was designed to be smaller than half the wavelength of the highest frequencies of the harmonics that were empirically shown to be relevant for the analysis, i.e., around 2000 Hz. In principle, a smaller spacing between microphones, a larger number of microphones, a larger overall array size, and full area coverage instead of only two lines can only be better, but the number of channels soon becomes unviable. The chosen microphone configuration thus represents a practicable number of channels, while still providing a sufficient overhead. The relevant criterion is always the reproducibility of the results within the tolerance of playing. Importantly, we have ensured that the coordinates remain within this tolerance when the microphone lines are rotated by 45° from the horizontal/vertical orientation shown in Figure 2, i.e., from the + orientation to the × orientation. Furthermore, if only every other channel is kept, the change in results is still small, but this is already outside the safe range and is not recommended, especially probably for other instruments with higher pitch.

The recording protocol includes the geometric relationships between the instrument, chair, and microphones, indicated in Figure 2. Air temperature and humidity were monitored during the recording sessions. The instruments were tuned to standard pitch 443 Hz in just fifths as usual (subsequent frequency ratios of 2/3 from the highest (A) to the lowest (C) string), but with an electronic tuner for safety; the tuning was monitored and readjusted as needed.

5 Results of the pilot study

In the present pilot study with 9 violoncelli, we analyzed the harmonic timbres of the four open strings of the cello. Each instrument is thus characterized by four timbres, one for each string. The cellist had to bow each open string 5 times in succession at a comfortable (mezzo)forte. Besides statistical improvement, the main purpose of the 5 repetitions was to measure the base tolerance, i.e., the natural tolerance of playing, which is an estimate of the uncertainty of the resulting harmonic timbre coordinates.

The transformation procedure of the recorded sound into harmonic vectors \( \mathbf{h} \) is summarized in Figure 1. The SVD analysis equation (1) is done for each string separately with the complete collection (all cellos, all repetitions) of harmonic vectors \( \mathbf{h} \) of that string.

The results are comprehensively presented in Figure 3. For each string, these are the harmonic timbre space, spanned by the principal timbre vectors shown in Figure 3b, and the corresponding harmonic timbre coordinates

⁴ It is common practice to further smooth the signal obtained from a diffuse sound field by averaging its spectrum over frequency bands. In the case of musical instruments, whose tones consist of narrow and stable frequency peaks, this is not possible. Even in a large concert hall one can hear how individual harmonics of a distant instrument playing a steady note become weaker and louder when the listening position (or the position of the instrument) is changed slightly. Only spatial averaging can help here.

⁵ We use 33 phase-matched 1/4-inch microphones (B&K type 4958). The microphones are declared with a frequency response in the range of ±2 dB from 50 to 10 000 Hz. Data acquisition was performed with 24 bits and a sampling frequency of 65 536 Hz (B&K Lan-XI data acquisition system).
Figure 3. (a) Three-dimensional harmonic timbre spaces of cello strings (animated versions: Supplementary Material, Movies S1 to S4). The small circles represent the harmonic timbres of 5 repeated notes per instrument, while the radii of the mean-centered spheres are their standard deviations, indicating the uncertainty of the playing. The projections of the spheres onto the coordinate planes help to locate the actual positions in space and show the exact values of the three coordinates. (b) The orthonormal basis vectors of the harmonic timbre space (the columns of $\mathbf{U}$) are the principal timbres resulting from the analysis together with the harmonic timbre coordinates of the instruments. Their components indicate increased/decreased levels of the harmonics relative to their mean levels in the collection. $S_i$ (in units of bel) are the importance (amplification) factors of the principal timbres, equations (1)–(3).
harmonic space is a linear vector space in which usual linear algebraic operations such as addition, scaling, rotation, and projection of harmonic timbre vectors can be performed by definition. These operations have corresponding effects on sound and are reflected in traceable changes in the perceived timbre. It is important to note that harmonic vectors exhibit a strong correlation with their perceived timbres, yet they are not synonymous with the perceived timbres. In algebraic language, harmonic timbre space is not homomorphic to a hypothetical perceived timbre space, as any mapping from physical reality to auditory sensation is necessary nonlinear, to say the least. For example, adding a harmonic timbre vector to a sound with a given harmonic vector will only approximately change the perceived timbre of this sound in the direction and for the magnitude of the perceived timbre of the added harmonic vector. However, since there is no definition of something like a perceived timbre space, and we do not now what a change in the direction of some perceived timbre would possibly mean, the simple well-defined algebra in the harmonic timbre space is the best approximation of those notions we have. It is, therefore, sensible to listen to the effect of an algebraic operation in the harmonic timbre space that transforms a harmonic vector $h^a$ to another harmonic vector $h^b$, and observe whether the resulting perceived timbre is indeed related to the perceived harmonic timbre of an existing sound with harmonic vector $h^b$.

To create such sounds, we are not thinking of pure synthesis from harmonic vectors, as this would certainly sound completely synthetic. Rather, the harmonic timbre can be analyzed and manipulated as an isolated feature of the original sound and then – through appropriate Fourier filtering (colorization, in the following) – applied to that sound so that all its other (nonharmonic, temporal) timbral features are preserved. This is an advantage of spectral quantifiers and is not possible with general descriptors, which are arbitrarily elaborate, generally nonlinear properties and are used only one way – to extract features from sound, but not to apply them to sound.

We now present sound examples related to Figure 3 that illustrate and concretise the discussed aspects. Currently, these examples also function as demonstrations, providing readers with a tangible opportunity to form their own opinions.

**Sound of principal timbres.** The sound clip (1) is the mean harmonic timbre (i.e., the origin of the harmonic timbre space) of the C string of the examined violoncello, represented by an actual, original sample, that is closest to the mean among all channels/repetitions/instruments. Then, to hear the principal timbres, it is colored in the direction of

---

6 Note that the sign of a principal timbre vector is arbitrary – if it is changed, the corresponding coordinate also changes sign. In the last example, the 5th overtone can also be strong – it all depends on the sign of the coordinate.

7 In our case, this means amplification or attenuation of spectral components in frequency intervals around harmonic peaks, corresponding to the integration intervals used to obtain harmonic vectors.

8 The audio examples are best listened to with high quality headphones, comparing two or more sounds in succession to most clearly perceive differences and similarities in timbre.
• negative first principal timbre \(1\) (1a) and positive first principal timbre \(1\) (1b). The sound of cello 5 (1c) with a large negative coordinate \(c_1\), Figure 3 (C string), is to be compared to 1a, and the sound of cello 4 (1d) with a large positive coordinate \(c_1\) to 1b.

• negative second principal timbre \(2\) (2a) and positive second principal timbre \(2\) (2b). The sound of cello 1 (2c) with a large negative coordinate \(c_2\), Figure 3 (C string), is to be compared to 2a, and the sound of cello 6 (2d) with a large positive coordinate \(c_2\) to 2b.

Relevance of harmonic timbre metric distance. In Figure 3, we see that the A strings of celli 2, 3, and 6 happen to have similar first three coordinates. This gives us the opportunity to check whether their audible timbres are really closely related compared to those of the other celli. The corresponding sound samples of cello 2 (3b), cello 3 (3c), and cello 6 (3f) confirm that this is indeed the case – their timbres are similar. The timbres of celli 1 (3a), 4 (3d), 5 (3e), 7 (3g), 8 (3h), and 9 (3i), whose first three coordinates are farther away, are all clearly different from celli 2, 3, and 6. In addition, the timbres of celli 1 and 7 with almost coinciding coordinates \(c_3\) and \(c_2\) and different coordinates \(c_3\) actually seem to be more closely related.

This was to demonstrate the auditory relevance of the metric distance in harmonic timbre space and already in its three-dimensional (3D) subspace – instruments whose coordinates are close together sound similar, while instruments with coordinates farther apart sound very different. That two instruments with similar coordinates in the subspace actually sound similar is not self-evident. It underscores the dominant importance of this subspace and a lesser importance of other dimensions in which the coordinates of the two instruments are generally different. Also not self-evident is that instruments with very different coordinates actually sound very differently. This speaks for the perceptual relevance of harmonic timbre coordinates, indicating a close connection between the harmonic timbre vector space and the hypothetical perceived timbre space mentioned above. If they specified perceptually neutral features (something that is measured but not heard), they could be very different, while the perceived timbres would still be similar.

Direction in harmonic timbre space and additivity of harmonic timbres. Here we show that a displacement in harmonic timbre space for a relative harmonic timbre vector, that is, a shift from a timbre with some coordinates to a timbre with other coordinates, has a well-perceived auditory sense.

First we target a timbre with arbitrary coordinates, starting from the origin (the mean harmonic timbre) of the harmonic timbre space in Figure 3, C string, and moving to the point with chosen coordinates. That is, we color the representative of the mean (the same as sound sample 1 above) according to the desired target. For demonstration purpose we set the target to coordinates \((-0.29, -0.13, +0.06)\) corresponding to cello 5/repetition 1, so that the colorization result can be directly compared to the actual sample of cello 5/repetition 1. The mean harmonic timbre with coordinates zero: the original sample of cello 5 selected as the target: (4b). The result of the colorization from coordinates zero to the target, using only the harmonic timbre subspace spanned by the first three principal timbres: (4c). This is to demonstrate the auditory relevance of direction in harmonic timbre space/subspace and to assess the adequacy of its 3D subspace in general.

Now we target a timbre with arbitrary coordinates, starting from a sample with any coordinates (and not necessarily from the origin as in the previous case) and coloring it accordingly. We select the sample of cello 1 (5a) as the starting point. The target coordinates correspond to cello 5 (5b) as before. The result of the colorization from cello 1 to the coordinates of cello 5, using only the harmonic timbre subspace spanned by the first three principal timbres: (5c). This is to demonstrate the auditory relevance of the additivity of harmonic timbres and to assess the adequacy of the 3D timbre subspace in general.

6 Discussion and conclusion

The close mutual connection of sound and harmonic timbre coordinates – their direct two-way relationship – is the central essence of harmonic timbre coordinates. In conjunction with the linear harmonic timbre vector space created by the analysis, which allows the addition of harmonic timbres in the manner of the usual vector algebra, it brings many possibilities of virtual sound processing into the physical world, for example as a quantitative basis for the design and optimization of physical harmonic timbres in situ.

The harmonic timbre coordinates are robust – their determination is reproducible within the base tolerance. Pilot tests have shown that they are also independent of the measurement space, with a somewhat extended tolerance (of the order of the larger spheres in Fig. 3), provided the space is not too small – a typical chamber hall already proves to be sufficiently large. Further characterization of the influence of the measurement space and enhanced measures to mitigate it are necessary. Ongoing systematic studies aim to refine the definition of tolerances, and consideration is given to measurement strategies that could reduce the geometrical specificity of the recording protocol.

The crucial importance of spatial averaging is illustrated by Figure 4, which shows as an example the highly scattered A string coordinates of the individual microphone channels. It is symptomatic that the differences between channels can be larger than the differences between instruments. Sound samples of one and the same tone recorded by different microphones of the array, i.e. by the microphones number 1, 4, 9, 14, 18, \(21\), \(26\), \(30\), \(33\) as examples, confirm that the perceived differences are indeed large.

The sensitivity and reproducibility of the procedure is such that the smallest changes in the tuning of the
instruments are systematically reflected in proportionally small changes in the determined coordinates. Accurate tuning is therefore a prerequisite. At the same time, this means that the coordinates allow systematic detection of the dependence of harmonic timbre on the nuances of tuning, a phenomenon familiar to musicians. The inherent invariance of the method to pitch shifts is very advantageous in this respect.

The harmonic timbre space in Figure 3 is based on a small collection of instruments. Therefore, each of the instruments occupies a more or less unique place in it, with unique coordinates. As the collection grows larger, the timbres of the instruments will inevitably begin to overlap (the coordinates represent positions in the harmonic timbre space, not fingerprints of the instruments), which is particularly interesting from a data mining perspective. It is impossible to predict the nature of this overlapping and what, if anything, it might tell us about different instrument families. Will the overlapping be more or less stochastic, or will there perhaps be some form of non-random clustering that correlates with some other property that could be identified, e.g., size, model, type of bridge, age, type and origin of wood, storage conditions, frequency of use, etc.? Repeatable measurements and reproducible quantitative results (coordinates) are a prerequisite for addressing the question of correlation, even if it has only a greatly reduced harmonic aspect.

Specific regions of the harmonic timbre space can be identified by psychoacoustic judgments of sound from those regions or preferably by prominent, representative, famous instruments with those coordinates. In establishing global harmonic timbre standards for instrument families, it is therefore desirable to include instruments of the highest rank in addition to a sufficiently large and statistically representative collection.

The harmonic timbre coordinates are exactly independent of the amplitude of the signal, which is completely described by the zeroth coordinate. This is convenient when we want to study how the harmonic timbre of an instrument changes with the dynamics of playing (the entire dynamic range from pianissimo to fortissimo). Here we get a harmonic timbre for each dynamics, or a properly parametrized harmonic timbre line for the dynamic range. Such a change of timbre is an important means of artistic, emotional expression, to which we are very susceptible, since we know it primally from the human voice. Hypothetically, instruments may differ greatly among themselves in this respect.

Also of interest are average or appropriately defined combined harmonic timbres over broader pitch ranges and their variations from note to note. In the case of strings, possible harmonic timbre variations associated with very small pitch changes (quarter tone or less) may also be of great importance. Presumably, these can also vary greatly from one instrument to another and are very important for audibility and perception of vibrato.

We should not forget that the harmonic timbre coordinates introduced encode the harmonic content of an instrument’s sound, i.e., its harmonic timbre in the strict sense. Although this is undoubtedly a very important property,
it represents only one aspect of the overall sound properties. Other aspects such as noise content, coarser and finer temporal modulation (texture), tone onset (attack) exist alongside and independently of the harmonic timbre coordinates.

Finally, we would like to point out once again that harmonic timbre coordinates can and should of course in no way replace the actual listening of instruments. They can, however, narrow down the selection of instruments to be evaluated by listening to a desired region of the harmonic timbre space. In this way, instead of choosing among several more or less random instruments of selected quality (price), one could choose among the same number of instruments preselected on the basis of their harmonic timbre coordinates and desired harmonic timbre. The desired timbre can be, for example, that of a famous instrument, or an instrument that one simply likes. It can also be shaped by virtually optimising a given timbre to one’s own preferences (e.g., to improve undesirable characteristics of an existing instrument – a possibility of interest to instrument makers), or even be generated synthetically. Such a rational strategy, which in no way diminishes the unique role of human aesthetic sensibilities and preferences, would increase the likelihood of finding an instrument that is truly close to one’s soul.

Acknowledgments

R.P. gratefully acknowledges the European Commission for funding the InnoRenew project (grant agreement #739574) under the Horizon2020 Widespread-Teaming program and the Republic of Slovenia (investment funding from the Republic of Slovenia and the European Union from the European Regional Development Fund). We acknowledge financial support from the Slovenian Research and Innovation Agency, grants Z1-4388 – Toward better understanding the diffuse sound field (R.P.), J4-3087 – Engineered wood composites with enhanced impact sound insulation performance to improve human well being (R.P., D.S.), and 10-0035 infrastructural program (R.P.). D.S. thanks the Department of Physics of the Faculty and the Laboratory for Molecular Modeling of the Institute for their comprehensive understanding and support. The authors thank the Vrhnika Music School and the owners of the violoncelli for their kind hospitality, generosity, and cooperativeness.

Supplementary material

A supplementary material of this article is available at https://acta-acustica.edpsciences.org/10.1051/aacus/2023071/olm.

Movie S1: Animated Figure 3: three-dimensional (3D) harmonic timbre space of the cello’s C strings.

Movie S2: Animated Figure 3: 3D harmonic timbre space of the cello’s G strings.

Conflict of interest

The authors declare no conflict of interests.

Data availability statement

Data are available on request from the authors. The sound files associated with this article are available in Zenodo public repository, under the reference [59].

References


Cite this article as: Prislan R. Kržič U. & Svenšek D. 2024. Quantifying sound colour of musical instruments – precise harmonic timbre coordinates of like instruments. Acta Acustica, 8, 8.