DOI: 10.61782/fa.2023.0725

forumacusticum 2023

HOW GOOD ARE WE AT DISTINGUISHING MUSICAL TEMPERAMENTS?

Emmanuel Friot[1*] **Sabine Meunier**[1] **Jacques Chatron**[1]
[1] Aix Marseille Univ, CNRS, Centrale Marseille, LMA UMR 7031, Marseille, France

ABSTRACT

An off-the-shelf digital organ controlled by a MIDI sequence was used to obtain recordings of an excerpt of early music that differed only in the chosen musical temperament. At the same time, the recording of slow chromatic scales made it possible to accurately identify the fundamental frequencies of the organ notes, and these frequencies were in agreement with the theoretical values of the temperaments indicated by the manufacturer of the organ. An Odd One Out test was conducted with the musical recordings to assess the extent to which listeners differentiate between the equal temperament and five other more or less unequal temperaments. Three-sound series were presented, with equal temperament twice and unequal temperament once. The middle interval always contained the equal temperament recording, the listener was asked to indicate which sound was different from it. Over forty participants, including non-musicians and early music professionals, took this test. It turned out that participants overwhelmingly discriminated between equal and other temperaments, with discrimination rates increasing when temperaments were varied in the usual "musicological" order from least unequal to most unequal. Those who had the most difficulty distinguishing temperaments were non-musicians, but those who had the least were not early music professionals.

Keywords: *psychoacoustics, music perception, musical temperament*

**Corresponding author: friot@lma.cnrs-mrs.fr.*

1. INTRODUCTION

Musical temperament, i.e. the exact relative fundamental frequencies of the scale notes, has evolved widely over the centuries in western countries [1]. The choice of which temperament to use for a given instrument and piece of music is a musicological question that can be very complex. For example, figure 1 shows the manuscript header of Johann Sebastian Bach famous Well-Tempered Clavier. It was not until 250 years after his death that the frieze in the figure was recognized as a coding by Bach of the temperament he favored [2]: the number of turns in each loop indicates how much to shrink the musical intervals in the circle of fifths.

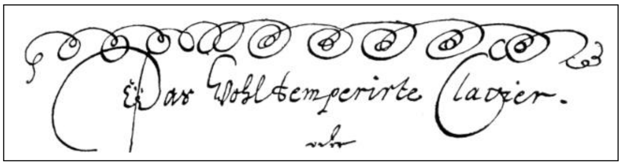

Figure 1. Bach's handwritten header for the Well-Tempered Clavier.

Tuning a viola da gamba or a harpsichord to change the temperament is possible during the intermission of a concert. It takes much longer in the case of an organ, for which (dozens of) years may pass from one general tuning to another. In this case the choice of a temperament, for example during the restoration of a listed organ, commits for a long time.

In order to assess the consequences of choosing one organ temperament over another, one might ask how well the audience hears the differences between temperaments. Numerous studies have been carried out on the perception of pitch and, in particular for the pedagogy of bowed string instruments, on that of musical tuning [3,4]. On the other hand, very few studies have focused on temperament perception, perhaps because it is not easy to obtain musicologically relevant excerpts for laboratory testing that differ only in temperament. In [5], a recording of a work by Nicola Vicentino was manipulated with auto-tune software in order to have one version in equal temperament and another in the particular temperament advocated by the composer. It was found that professional musicians could differentiate the two versions very well. However, adjusting the fundamental frequency of each note in a recording does not allow easy comparison of a larger number of temperaments and musical excerpts.

In this context, a modern digital organ, driven by a MIDI file, was used in the study presented here to generate musical excerpts differing only in temperament as set by the organ control interface. The digital organ controlled in this way produces musical samples in a perfectly reproducible way. An excerpt of Italian music from the 17th century was used to conduct tests to evaluate the extent to which listeners, from non-musicians to early music professionals, hear the differences between versions. At the same time, the recording of chromatic scales allowed to estimate precisely the playing frequencies of the organ in order to compare them with the expected theoretical values. The recording of the sound excerpts, the identification of the playing frequencies, an Odd One Out psychoacoustic test on the distinction of the temperaments and its first results are presented in more detail below.

2. RECORDING OF MUSICAL EXCERPTS

Figure 2 shows the off-the-shelf Johannus Classic 350 digital organ used for this study. This organ has 4 quality sound banks and a reverb using the impulse responses of 12 identified buildings. 11 temperaments are proposed by the organ menu, arranged in a usual musicological order from "Equal" (two consecutive notes are in a fundamental frequency ratio of $2^{\frac{1}{12}}$) to" Pythagorean" (the scale is built from consecutive perfect fifths). The organ has a MIDI interface and a stereo audio output. A RME Babyface PRO FS audio card and Adobe Audition software was used for this study to control the organ in MIDI and to record its audio output. The signals were recorded with sampling rate $f_s = 48$kHz and 24-bit resolution.

One cannot choose just any musical excerpt to evaluate the extent to which one perceives differences between musical temperaments. Most listeners can instantly hear a problem when a chord of F♯ major or A♭ major is placed on a keyboard in mesotonic 1/4 comma temperament, but

Figure 2. The LMA digital organ

no music has been written for this temperament in these keys. It is therefore necessary to choose musical excerpts and temperaments whose association corresponds to a real cultural practice. In order to make sure that the sound samples are all different by varying the temperament, it is also desirable that the musical excerpt contains as many notes of the scale as possible: very different temperaments can have several notes in common, like for example F and A in Kirnberger II and 1/4 comma mesotonic temperaments.

Figure 3 shows the score of an excerpt from the Toccata settima by Michelangelo Rossi. This Italian piece from the first half of the 17th century was written for the very unequal 1/4 comma mesotonic temperament, then in use, but versions of it can be found on recordings in the 21st century on equal temperament instruments by professional organists. The two bars in figure 3, which correspond to less than 10 seconds of music, include the 12 steps of the scale as usually present on a keyboard without split sharps (i.e. C♯, E♭, F♯, G♯, B♭ and not D♭, D♯, G♭, La♭, A♯). This musical excerpt therefore fulfills the required conditions and its duration is compatible with the usual psychoacoustic tests.

The excerpt shown in figure 3 has been recorded in the 11 temperaments available on the digital organ, knowing that one of them (the Pythagorean temperament) is older and not quite suitable to the excerpt. The MIDI file used to drive the organ was generated from the music editing software Musescore with tempo 80 bpm. The historic sound bank used for the recording and the reverb, set to maximum, were the one indicated by the organ manufacturer as

Figure 3. Bars 75-76 from the Toccata settima by Michelangelo Rossi.

those of the Oude Kerk in Amsterdam. The default pitch of the organ (A4 at 440Hz) was used. Finally, of all the possible choices of organ stops, the manufacturer's preconfigured *Plein-jeu* was selected, as the large number of harmonics then present may enhance the specificities of each temperament.

3. ESTIMATION OF THE DIGITAL ORGAN PLAYING FREQUENCIES

For this study, it was necessary to check that the playing frequencies of the digital organ match the expected theoretical values for every temperament. For this purpose, slow chromatic scales were recorded with the same organ settings as for the excerpt shown in figure 3, except that the reverberation was set to the minimum for a more accurate estimation of the playing frequencies. The estimation was done for each note of the central octave of the digital organ with a procedure, which has been used in churches to identify at several points the same playing frequency to within 0.01Hz (cf. [6]), which consists in:

1. low-pass filtering the signal with cut-off frequency about 1.5 times the expected playing frequency,

2. computing, with a Hilbert transform, the analytical signal associated with the quasi cosine signal obtained after low-pass filtering,

3. unwrapping phase ϕ of the analytical signal and computing the playing frequency at every time sample $f_s \frac{\phi_n - \phi_{n-1}}{2\pi}$,

4. estimating the mean and the standard deviation of the note playing frequency.

In the case of the digital organ, the standard deviations of the instantaneous frequencies appear to be one order of magnitude smaller than those measured on pipe organs in churches. The average frequencies are therefore probably estimated with an accuracy better than 0.01Hz. The playing frequencies identified for each of the two channels of the recorded stereo signal are identical to better than 0.001Hz.

A first comparison of the estimated playing frequencies of the digital organ with the expected theoretical values revealed differences that were much greater than the precision with which these frequencies are estimated. Therefore, several recordings and calculations were made for each note. Table 1 shows, for the case of equal temperament, the theoretical values obtained from 3 recordings. Significant differences between recordings can be seen, perhaps related to natural fluctuations with the temperature of devices such as the digital-to-analog converters of the organ. Although not negligible, these frequency variations are quite small compared to the theoretical differences between temperaments and certainly smaller than the tuning imperfections of any acoustic musical instrument. We can therefore consider that the temperaments proposed by the digital organ are in accordance with the theory.

Table 1. note frequencies in equal temperament: theoretical values and estimated values from 3 measurements at the numerical organ

note	th. fr.	est. fr.#1	est. fr.#2	est. fr.#3
C4	261.63	261.92	261.68	261.92
C♯4	277.18	276.86	277.37	277.62
D4	293.66	293.73	293.86	293.99
E♭4	311.13	311.33	311.33	310.77
E4	329.63	329.70	329.85	329.85
F4	349.23	349.78	349.78	348.82
F♯4	369.99	369.90	369.73	370.07
G4	392	391.72	392.44	392.26
G♯4	415.3	415.77	415.39	415.77
A4	440	440.09	440.09	440.49
B♭4	466.16	466.47	466.47	466.47
B4	493.88	494.66	494.21	493.31
C5	523.25	523.6	522.88	522.88

4. IMPLEMENTATION OF THE PSYCHOACOUSTIC TEST

A dissimilary test by pair comparison, with the sound samples obtained on the digital organ in all 11 temperaments except the Pythagorean one, was originally envisaged to estimate perceptual differences between all temperaments. Once the samples were obtained it seemed to us that such a test would be too difficult for the participants. An Odd One Out test was therefore preferred to assess the extent to which a listener differentiates between equal and unequal temperaments. The listener was presented with a series of 3 sounds, 2 of the 3 sounds corresponding to the musical excerpt in equal temperament and the first or the third to the one in unequal temperament. After listening to the 3 sounds, the listener had to indicate if the sound different from the other two seemed to be the first or the last. A feedback in the form of a green or red light then indicated if his choice was the right one or not and a new series of 3 sounds was played.

The chosen musical excerpt lasts about 10s including reverberation, the series of 3 last 30s. By presenting 10 times each excerpt in unequal temperament, we chose to carry out tests lasting 25 minutes by restricting ourselves to 5 of the 9 possible unequal temperaments: Young II, Valloti, Kirnberger II, Werckmeister III and mesotonic 1/4 comma. The test sessions began with a 5-minute familiarization phase during which 10 series of 3 sounds were played, with each of the unequal temperaments played once as the first and once as the last in the series. In this way, all the sounds of the corpus were presented to the listener before the actual test. Figure 4 shows the interface offered to the listeners during the test.

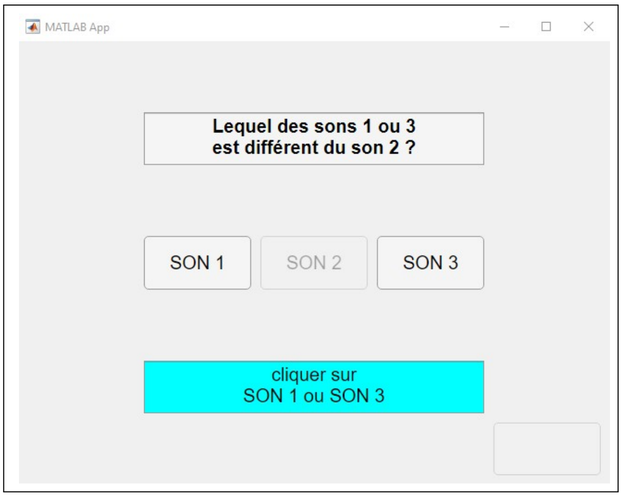

Figure 4. The psychoacoustic test interface.

5. FIRST RESULTS

45 people took part in the listening test. 6 of them had never had any musical practice or education outside of compulsory education. On the opposite, 8 of them were early music professionals: 6 harpsichordists, one baroque violinist and one singer. This sample of 45 listeners is certainly not representative of the general population, but it could constitute an early music concert audience.

Figure 5 displays the test results for the 45 listeners as a box plot. The dispersion between listeners is large, but 75% of the listeners have a rate of correct answers well above 50%, i.e. they did make the difference under laboratory conditions between equal and unequal temperaments even for only slightly unequal temperaments. 42 participants out of 45 reached 100% correct answers with the mesotonic temperament at 1/4 comma. The order in which the temperaments are arranged in figure 5 is the order of the digital organ menu, which is also the usual ”musicological” order in which these temperaments are arranged from least unequal to most unequal. The medians of the listener discrimination rate increases following this order, the musicological and perceptual rankings are therefore in agreement.

While waiting for a more detailed statistical analysis of the results of the test, it should be noted that the outliers shown in figure 5 correspond to test participants who do not play music at all. On the other hand, 5 people obtained 100% correct answers for all temperaments. None of these 5 people is an early music professional, but all of them had already taken psychoacoustic tests before, which was not the case for the early music professionals who took the test.

Figure 6 displays the test results for the 8 early music professionals. The results are more erratic than in figure 5 because of the small number of listeners but the median levels are similar. Moreover, the medians in figure 5 are not modified if these 8 professional musicians are removed from the panel of listeners.

6. CONCLUSION

A psychoacoustic Odd One Out test was set up to evaluate a listener's ability to distinguish between two recordings of a 17th century musical excerpt of less than 10 seconds

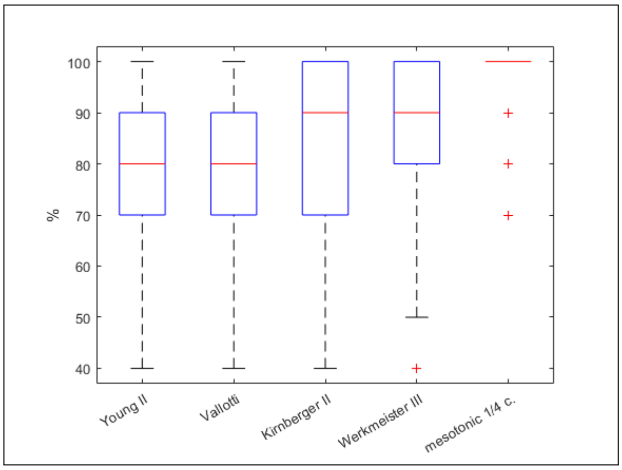

Figure 5. Boxplot of the psychoacoustic test results. An horizontal bar denotes a median, a rectangle goes from first to third quartile and the dashes cover from first to ninth decile. Red crosses indicate potential outliers.

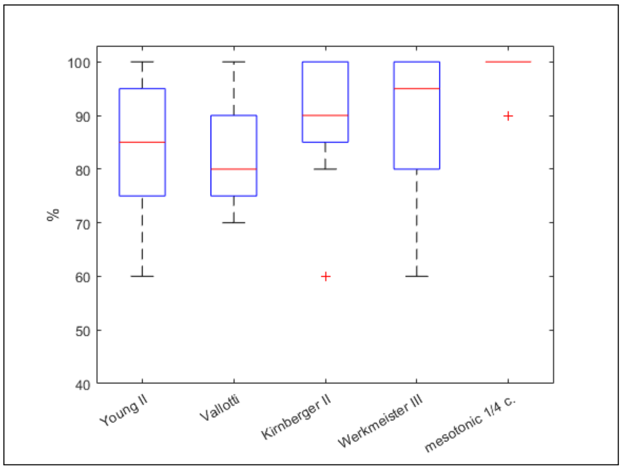

Figure 6. Boxplot of the psychoacoustic test results for the 8 early music professionals.

duration, one in equal temperament and the other in unequal temperament. The sounds were recorded at the output of a digital organ driven by a MIDI file, only the temperament differs between them. 45 participants, with very different musical skills, took a test using 5 unequal temperaments. It turned out that the participants distinguished very well between equal and unequal temperaments. This result is similar to the one presented in [5] for the case of equal temperament and the one recommended by Nicolas Vicentino. It was also found here that the discrimination rates increased when temperaments varied in the usual "musicological" order, from least unequal to most unequal. The people who had the most difficulty distinguishing temperaments were non-musicians, but the 5 people who achieved a 100% discrimination rate were not early music professionals.

Now that we know that equal and unequal temperaments are easily distinguishable, we could construct a perceptual space to see what are the respective perceptual distances between a large set of temperaments. To do this, we will soon carry out a dissimilarity test using pairwise comparisons of musical excerpts.

7. ACKNOWLEDGMENTS

Many thanks to all the musicians who contributed to this work and especially to the early music ensemble La Camerata Chromatica.

8. REFERENCES

[1] S. Isacoff, *Temperament: How Music Became a Battleground for the Great Minds of Western Civilisation*. Faber & Faber, 2007.

[2] B. Lehman, "Bach's extraordinary temperament: Our Rosetta Stone–1," *Early Music*, vol. 33, pp. 3–24, Feb. 2005.

[3] D. E. Hall and J. T. Hess, "Perception of Musical Interval Tuning," *Music Perception*, vol. 2, pp. 166–195, Dec. 1984.

[4] M. P. Lynch and R. E. Eilers, "A study of perceptual development for musical tuning," *Perception & Psychophysics*, vol. 52, pp. 599–608, Nov. 1992.

[5] M. Miller, "Nicola Vicentino and the Enharmonic Diesis: An Analytical and Empirical Study," Master's thesis, McGill University, Montréal, Québec, Canada, 2011.

[6] E. Friot and J. Chatron, "Effet de la durée des notes sur le niveau sonore d'un orgue," in *16ème Congrès Français d'Acoustique*, Marseille, 2022.