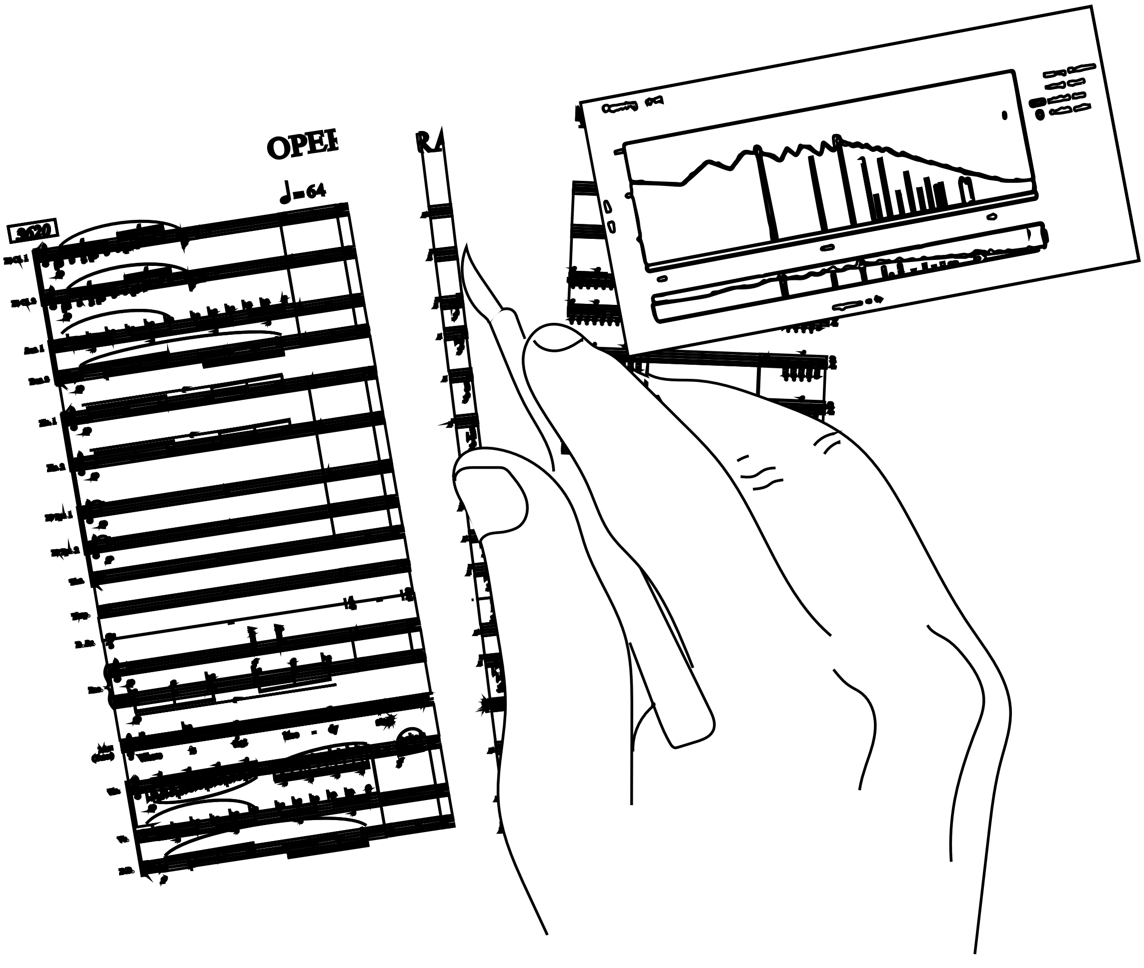
SCORE-TOOL



TARGET AUDIBILITY TESTING IN ORCHESTRATION

Doctoral project by Uljas Pulkkis

EST 72

Sibelius Academy of the University of the Arts Helsinki 2023

Supervisor:
Marcus Castrén, Prof. Emeritus – Sibelius Academy, University of the Arts Helsinki

Pre-examiners:
Professor Stephen McAdams, PhD, DSc, FRSC – Schulich School of Music, McGill University, Canada
Composer Jean-Baptiste Barrière, PhD from Université de Paris I Panthéon-Sorbonne, France

Chair/Custos:
University Lecturer Ulla Pohjannoro, DMus – Sibelius Academy, University of the Arts Helsinki

Examiner:
Professor Stephen McAdams, PhD, DSc, FRSC – Schulich School of Music, McGill University, Canada

Score-Tool – Target Audibility Testing in Orchestration

EST 72

Applied Study Programme (Artistic Degree)

DocMus Doctoral School

Sibelius Academy of the University of the Arts Helsinki

Cover Satu Grönlund

Cover image Uljas Pulkkis

Printed by Hansaprint Oy

ISSN (PDF) 2489-7981

ISSN (print) 1237-4229

ISBN (PDF) 978-952-329-301-4

ISBN (print) 978-952-329-309-0

Helsinki, 2023

Dedication

This project is dedicated to the memory of my beloved parents, Anneli and Göran Pulkkis, who both passed away in 2021.

Acknowledgements

I would like to thank my supervisor Marcus Castrén for his dedicated support and guidance. A big thanks goes to Tapio Lokki too, who continuously provided encouragement and was always enthusiastic and willing to assist in any way he could throughout the doctoral project. I would also like to thank Tuukka Ilomäki for providing advice regarding computer programming, and Andrew Bentley for encouragement and valuable insights on technical matters. Ulla Pohjannoro took care of the University's practical matters. Glenda Dawn Goss wrote the libretto for the opera that is a part of this project.

A special thanks goes to the soloists of my opera *All the Truths We Cannot See* and the Sibelius Academy opera department along with the USC Thornton School of Music opera who produced it.

Thanks also to my wife Anna, for putting up with me sitting in the front of the laptop for hours, and for providing guidance and a sounding board when required. You have no idea how much happiness you have brought into my life. For my children Elsa and Uuno, I'm sorry for being even grumpier than usual while I worked on this project. A special thank you for my parents, Göran and Anneli who always supported my projects and goals; it is hard to forget you who gave me so much to remember. Finally, many thanks to all participants who took part in the study and made this project possible.

Abstract

This is an applied artistic doctoral project, which is divided into three parts: an orchestral composition, a computer application and its source code, and a written report. In this project, I present an orchestration issue that I have encountered in rehearsals of my compositions. The issue is the inaudibility of an instrument or an instrumental group that I thought to be audible when I was writing the score. This inaudibility can occur even though the orchestration follows the best practices recommended in academic handbooks. The artistic part of the project is an opera *All the Truths We Cannot See*, which I composed while acquiring new knowledge about orchestration as described in this report.

The developmental part of the project is the Score-Tool computer application, Score-Tool App, which I designed and coded myself. The application is intended for composers to enable them to analyze a score during the composing stage in terms of the audibility of a desired instrumental sound. The audibility is calculated based on real instrument audio analysis and measurements of sound intensity levels in a performance situation. In addition to my own research, the App utilizes various psychoacoustic algorithms, borrowed from lossy audio coding- and speech recognition applications and virtual fundamental research to determine the masking threshold of the orchestration. I have also adjusted the existing algorithms for this project's purposes. Any sound that has spectral peaks above the masking threshold within at least one *critical band* should thus be at least partly audible. The inaudibility of a sound can also be caused by its *auditory blending* into the orchestration. The blending of a sound can happen, for example, when its timbre matches the orchestration or when the *spectral centroid* of the sound is low in frequency space. The application estimates the blending of the sound by calculating its spectral centroid and by comparing the timbre similarity with orchestration by utilizing the *MFCC* algorithm, which is borrowed from speech recognition applications. As a result, the App calculates the *audibility prediction* in percentage values. The audibility prediction is my own term, calculated with the algorithm I developed. This value comes from a combination of masking and blending algorithms whose importance is weighted based on my experiences and my own measurements in orchestral rehearsals. The usability of the Score-Tool App has been tested by me and by other composers by analyzing both existing and in-progress compositions using the App.

The report is divided into three parts. In Part I discuss aspects affecting the sound audibility in orchestration, what is taught about the subject in music universities, my personal experience in this area, the effect of the hall and the performer's position on stage, and how the issue is addressed in music psychology research.

In Part II I describe the development of the Score-Tool App, how the algorithms work "under the hood," and the features I implemented in the program. Part II also includes my own research in visualizing orchestration masking and timbre data as well as a tutorial for the App. The App manual can be found in the tool itself. The Score-Tool App is also addressed at the end of the report, where I discuss the possibilities for future research in this field using the Score-Tool App and other orchestration features that could be implemented into it.

In Part III of this report, I describe the testing phase and how using this App has changed the way I write for the orchestra while composing the opera *All the Truths We Cannot See*. In addition, I present cases where this App has helped my colleagues in their orchestrations. I also discuss how the analysis results of the Score-Tool App correlate with live performance and how the audibility prediction algorithm of the Score-Tool App has been adjusted based on my

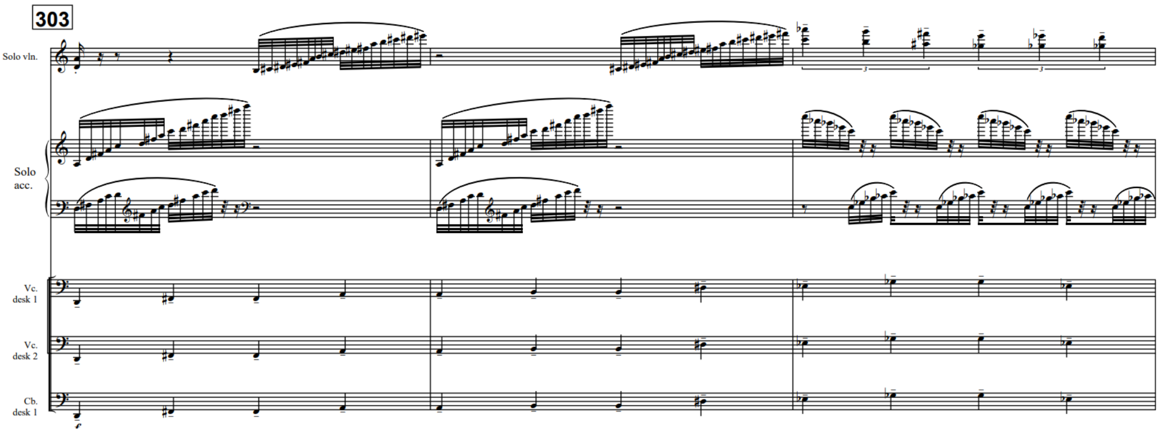
experience and measurements I made in this testing phase. In other words, I test the validity of the results that the Score-Tool App currently gives a composer.

The conclusion is that the Score-Tool App provides relevant information about possible audibility issues in a composition. Using the App does not solve the issues immediately, but it does give reliable specifics about which instruments are causing the issues. This information helps composers during the artistic process to pre-evaluate the functionality of their orchestration. The App guides the composer in a field that, owing to its complexity, has very few definite answers.

TABLE OF CONTENTS

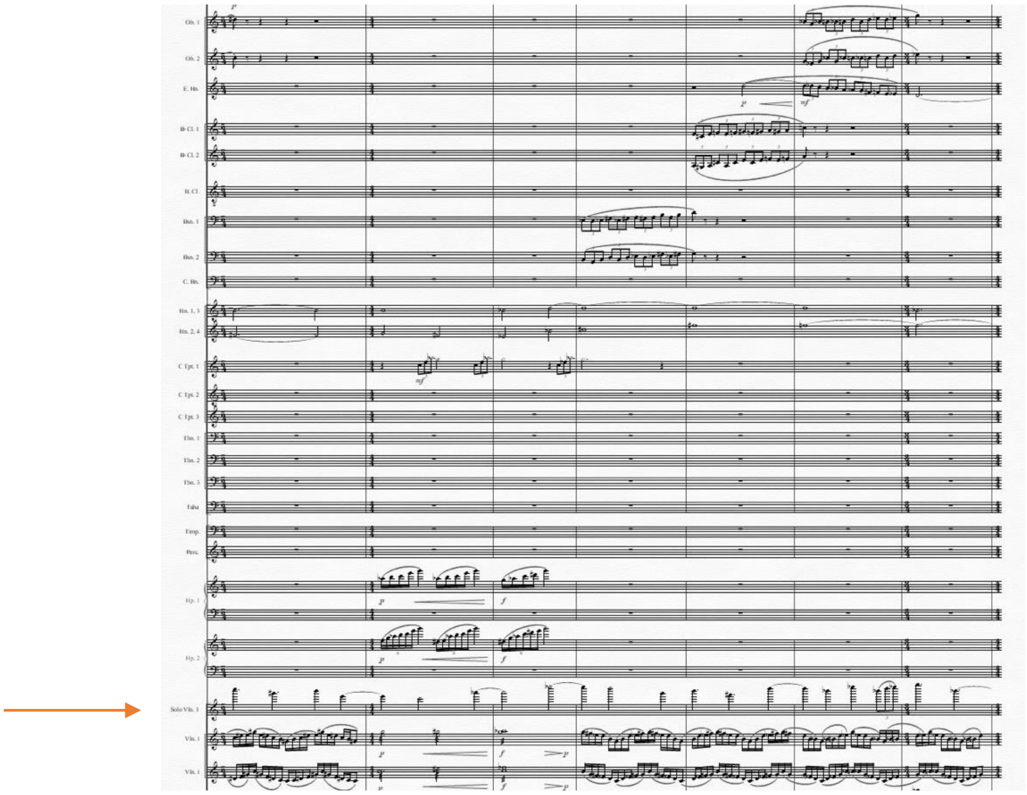
INTRODUCTION

MOTIVATION

The initial idea for this doctoral project came from my experiences in orchestral rehearsals of my compositions. In those rehearsals, especially in the first readings of a score, I encountered the same issue several times: an inaudible instrument sound. I do not expect to hear every single instrument in the orchestra all the time, but as a composer, I have certain audibility preferences in mind when I am composing a piece. The balance of the orchestral texture may not sound as I intended, either because of the choice of dynamic markings, the registers, the playing technique, or instrumentation. The issues arise especially when I use non-conventional instrumental combinations and textures. The audibility matter is not the result of one's inexperience in orchestral writing, but rather is a side-effect of a composer's artistic choices in the complex orchestral-apparatus. Sometimes even when everything is done as taught in academic orchestration handbooks, the main thing in the orchestral texture, the one thing I had in mind at the composing stage can be heard only faintly or is completely drowned out in the mass.

Example 1. Pulkkis, *Crystallizations*, a double concerto for accordion, violin and string orchestra, mm. 303-305. The solo accordion is inaudible in these bars with the exception of its very highest notes. The dynamic marking for accordion and double bass is *f*, and dynamic marking for cellos is *mf*.

Three examples from my compositions can demonstrate the issue. First, in the rehearsals of my double concerto for accordion and violin, *Crystallizations*, there were numerous passages where the sound of the accordion was masked by the orchestra. This surprised me because I intentionally paid careful attention to the audibility of the solo instrument. I placed the accordion mainly in its own register in frequency space, marked it with louder dynamics than the orchestra's, and specified the loudest registration. In the rehearsals, lowering the dynamics of the instruments playing in nearby octaves did not help very much. In one passage, the conductor tried to balance the texture according to my wishes, and the guilty instrument appeared to be the double bass playing two octaves below in register. Because of limited rehearsal time, a similar balancing of texture was not carried out in other passages.

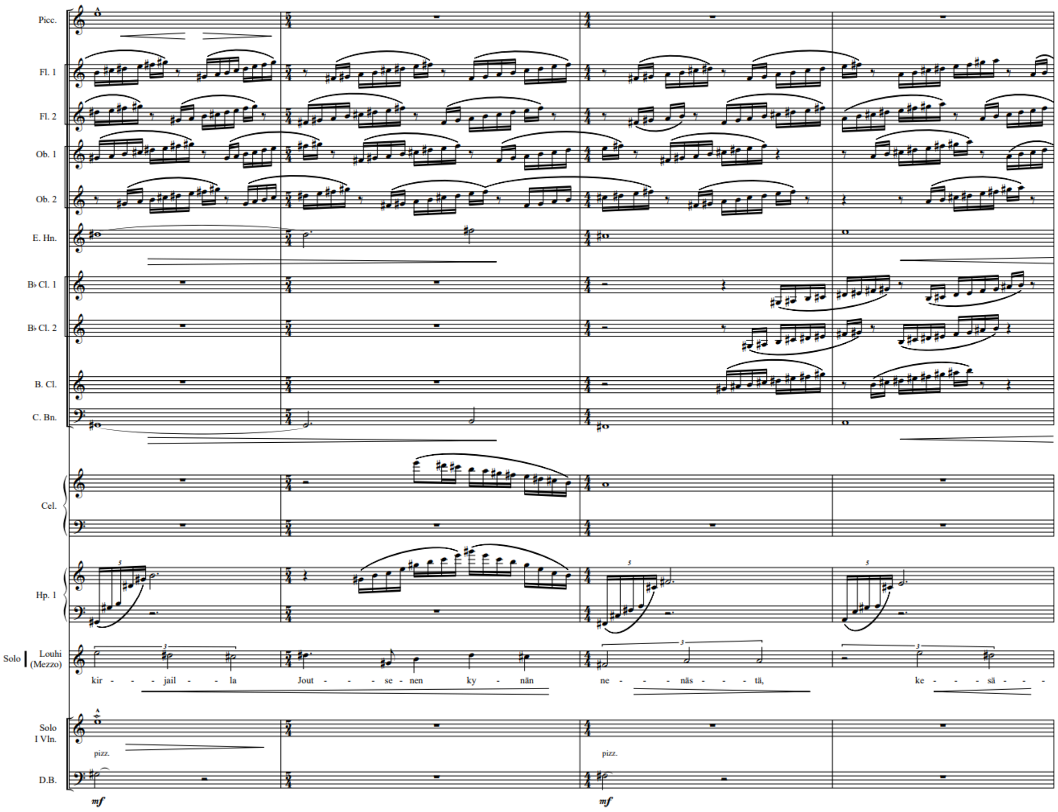
Another surprising audibility issue occurred in my orchestral work *Trial*. The piece does not call for any one soloist to stand in front of the orchestra, but I have written solos for instruments in individual passages. I composed the piece for the Sibelius Conducting Competition. A world-class conductor read through the score before the performance to check that there were no technical problems in the orchestration. A surprise came for all competitors in the passage shown in Example 2. Here a violin solo, composed in register two octaves above the nearest instrument and played by concertmaster was nearly inaudible, even at the conductor's podium beside which the soloist was playing. Even lowering the dynamics in the orchestra did not help that much, so the problem was in the orchestration. For me, the biggest surprise was that even the highly experienced conductor did not spot the problem in the score. In a later version, I corrected the audibility of the passage by doubling the soloist with two more solo violins an octave below.



Example 2. Pulkkis, *Trial* for symphony orchestra, mm. 61-67. The solo violin (marked with an arrow) is inaudible in this passage, even in *forte* dynamics. I revised the passage later by doubling the violin with two more solo violins an octave below.

The third example is from my operatic cantata *Neito* ("*Maiden"*), which I composed for orchestra, soprano, mezzo-soprano, and male choir. I used a wide palette of orchestral timbres, including in passages with female soloists. In the rehearsals, there were numerous places where the voices of both solo singers were unexpectedly drowned out, even by orchestration with low dynamics. By drowning, I mean that not only were the voices inaudible, but also the intelligibility of the text suffered, and the expressive beauty of the voices could not be heard.

This was understandably a problem for soloists as well because they did not want to sing in in *fortissimo* too long in order to save their voices. The cantata is an hour in length, so not much rehearsal time could be spent on balancing all the problematic passages. The conductor simply gave the orchestral players a general rule: in passages with a soloist, play very softly. I ended up even cutting some parts in passages with unsolvable problems.

The following example (Example 3) is one of the problematic places. Although flutes and oboes are playing in the same register as the voice, these are not the instruments drowning the soloist. In the rehearsals, removing the double bass and the double bassoon temporarily helped, but that revision made the orchestra sound thin and unattractive. So I kept the orchestration as is, but the mezzo soprano tried to sing louder.

Example 3. Pulkkis, *Neito*, operatic cantata, mm. 204-207. The double bass and double bassoon, surprisingly, turned out to disturb the vocal line.

All three cases show that the usual fix is to lower the dynamics of the orchestration and raise the dynamics of the soloist. In my experience, this gives unsatisfying results, because lowering the dynamics of the orchestra leads to lack of intensity in a passage, while raising the dynamics of the soloist leads to lack of expression. I started the Score-Tool project to learn and understand the functionality of complex orchestration and to learn how to orchestrate intense passages even when the texture contains solo instruments.

THE PURPOSE OF THE PROJECT

As a composer, I find that orchestration is an integral part of my composition practice. That means that I do not first compose a piece, for example, for piano and then orchestrate it later. Instead, I compose directly on orchestral staves and choose the timbres and textures I want to use, often intuitively. In other words, this practice does not allow a computer-generated orchestration because I do not orchestrate, but rather compose directly for orchestra.

In seeking the right solution for the balance problems explained above, I came to the conclusion that I needed a tool to tell me if my score has audibility problems, a tool that does not orchestrate for me, but tells me if my intuition has gone wrong. I came up with an orchestration analysis tool, a computer program, to test the audibility of the "main thing" in the score. Thinking further about my needs, I realized that, in my composition, often the "main thing" is a combination of instruments or a complex texture that I want to be audible, not as a cluster of individual timbres, but as one entity. In this project, the "main thing" is called the *target.* The target can thus be the sound of one instrument, for example, a solo violin in a violin concerto or part of flute section in an orchestra or a combination of instruments, in which case there is the question of timbre homogeneity.

All these questions of audibility and homogeneity lead to the subject of psychoacoustic analysis. Target audibility is not a property of orchestration that is on or off; all the frequencies are there in sounding music, even the ones that our hearing system cannot detect. Therefore, I would need to use psychoacoustic methods developed especially to mimic the human perception of sound, in this case, the sounding orchestration. In our digital era, the natural choice for a psychoacoustic testing tool for orchestration is a computer application. In this project, I have developed methods to apply psychoacoustic algorithms to orchestration analysis to obtain greater knowledge of orchestral writing.

Because this is an artistic doctoral project, the interesting question I posed for myself was this: Does knowledge of psychoacoustic properties of orchestration change the way I compose and orchestrate music? Knowing the potentially good and bad combinations of instruments may lead a composer to favor the good and avoid the bad. In the concluding chapter, I discuss this issue from the perspective of my opera *All the Truths We Cannot See*. This whole project is a work from the perspective of an orchestral composer, me, and therefore, the discussion in this report about pros and cons of the Score-Tool App is purely from an orchestral composer's perspective.

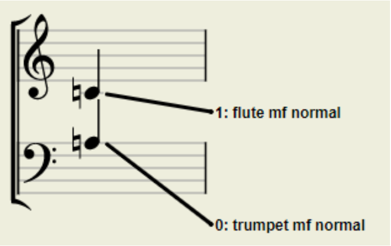
BENEFITS OF COMPUTATIONAL AUDIBILITY TESTING

Testing the audibility of the target should not be limited to composers only. In my discussions with conductors, many revealed that there are numerous works in the standard orchestral repertoire which have major balance problems. Problematic examples include concertos, symphonies, and operas in which a score does not sound well with the composer's specified dynamic markings or even in the register in which a soloist's part is written. Audibility testing can thus be used, not only by composers, but also by conductors or instrumentalists to pre-evaluate a score before an orchestral rehearsal. With pre-evaluation, it is possible to detect passages that need special attention and perhaps even make revisions beforehand to save rehearsal time. Audibility testing also identifies the exact instruments which compete with the soloist; lowering just those dynamics may solve the audibility issues without changing the dynamics of the whole orchestra.

In addition to orchestral rehearsal situations, there seems to be a huge amount of tacit knowledge among conductors about how to handle the balance of certain well-known works. Conductors can also, at will, check if the tacit knowledge about the balance in well-known works correlates with the results of the pre-evaluation. However, the pre-evaluation cannot give absolute dynamic values that would make the target audible, but they can give a rough estimate of the audibility.

TRIVIAL AND NON-TRIVIAL AUDIBILITY ISSUES

Audibility issues may be trivial in cases where a weak-sounding instrument is playing with a strong-sounding instrument in approximately same the register. An example is shown below where the flute sound will probably have audibility problems.

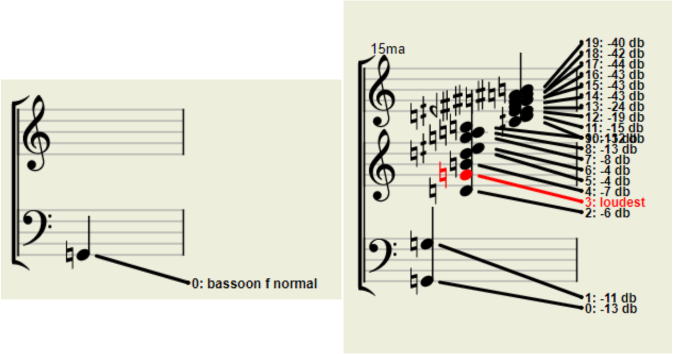


Example 4. The flute sound will probably have audibility problems because of the strong sounding trumpet in the same register. The word "normal" refers to normal playing technique.

Both instruments in Example 4 have the strongest sounding harmonic at written pitch, which makes the detection of audibility issues intuitive. However, if there are instruments in the orchestration that do not have the strongest harmonic at their notated pitch, then audibility issues might come as surprise.

In particular, some low-register instruments may have the strongest sounding harmonic up to two octaves above the written pitch. In Example 5, the pitch *g2* is notated for bassoon along with its sinusoidal partials with their relative loudness, the strongest partial being marked 0 dB. The example shows how the loudness of the notated partial, *g2*, is down to 13 dB softer than a partial *g4* two octaves above. A weak-sounding instrument notated two octaves above the written pitch of the bassoon might thus be masked, even though the instruments are not playing in the same register.

Audibility issues with complex orchestration can become even more unpredictable than in Example 5 . For example, in a 10-part orchestration with variable dynamics, spread widely in register, there is no intuitive way to predict whether the intended target will be audible or not.



Example 5. The bassoon playing the notated pitch *g2* produces the strongest sounding partial two octaves above the written pitch.

HOW CAN THE AUDIBILITY OF A TARGET BE TESTED?

Target audibility testing would be a perfect feature for a notation program, but for that purpose, the app would be bound to a commercial product, which is not my intention. Thus, the testing must be done on a score that is non-commercial in format. One such format is MusicXML, an open format that is readable and usable with a wide range of music notation applications. This is the format I selected to use for my project.

For orchestral compositions written with some of the familiar notation programs, such as *Dorico*, *Finale*, and *Sibelius*, scores can be saved in a MusicXML format directly from the notation program. For pre-computer era music, the score can often be converted from a PDF or photo to MusicXML format, for example, with non-commercial open-source tools such as *Audiveris*[1], although errors converted orchestral scores need to be corrected by knowledgeable humans.[2]

In the computer application that is part of my doctoral project, the orchestration data are extracted from the digital score by mapping the pitches and dynamics of each instrument's notes on an orchestral instrument sound analysis database, created as a part of this project based on The University of Iowa Musical Instrument Samples (MIS) and an open-source orchestral library (VSCO orchestra). which contains detailed data on each instrument's acoustical properties as these sound in a concert hall. With the mapped data, it is possible to produce an acoustical model of the orchestration without actually performing the piece with a live orchestra. The acoustical model can be further analyzed with different algorithms.

For analyzing the perception of the target and the orchestration sounds, the acoustical model is put through the mathematical model of our hearing system. Combining the analytical data of each instrument in orchestration, calculating the psychoacoustic masking threshold produced by them, then comparing that threshold against the analytical data of the target, it is possible to predict whether or not the target has audible spectral components in its sound. The concept of psychoacoustic masking is defined in chapter "As mentioned earlier, when the basilar membrane envelope resulting from the spectrum of a more intense sound includes the envelope resulting from a less intense sound. If the amplitude of the traveling wave is large, it activates a cluster of hair cells creating a spread. So, the amplitude affects the width of the spread. This is the direct consequence of the basilar member's resonance. Loud amplitude creates bigger resonance, which affects a wider area.

The number of critical bands in an individual hearing system is unique, but there are several well-known, good approximations that are used in acoustic applications. In my App, I use Bark bands, which divide the frequency range of our hearing system into 24 critical bands. This model is a good representation of critical bands discounting borderline cases. A borderline case is when a frequency is situated close between bands.

In my Score-Tool App, I have used the MPEG psychoacoustic model, which divides the hearing range into 108 bands, the fractions of the Bark bands. The App keeps track of which sub-band belongs to which Bark band. Thus, when using a 108-band model, one must keep in mind that one band does not equal a critical band, but only approximately one-fourth of it.

Auditory masking".

With the analytical data in hand, the homogeneity of the timbre can be estimated. This is valuable information for composers who want the texture to sound as one entity and not as a collection of individual instruments. Also, in rare cases, very homogeneous orchestration can

---

[1] GitHub - Audiveris/audiveris: Latest generation of Audiveris OMR engine

[2] Daigle 2020

cause the target sound to be inaudible. In those cases, the target sound is not masked, but rather blended into the orchestration. The blending can also happen when the target has a low spectral centroid value, which is calculated by the program. A trivial case of blending would be scoring a flute solo accompanied by a pipe organ with flute registration, which at matching dynamics would be very close to each other in timbre.

I have combined and coded the instrument database and the algorithms in a computer application called the Score-Tool App. For the App, I have developed a method to calculate the audibility prediction value for the target sound as well as methods to visualize the psychoacoustical properties of orchestration timbre. The App is introduced in Part II, Chapter 1.

OBJECTIVES

PERSONAL OBJECTIVES

With the help of the Score-Tool App, I can pre-evaluate a score to reduce negative surprises in orchestral rehearsals. With this help, I can get feedback on the functionality of my compositional choices, which can further lead to calibrated choices that would match my initial intentions. A hypothetical situation would be a passage for a singer and orchestra, where the intention is to have a strong and big-sounding expression. This could lead me to write orchestration too thick and with loud instruments in a register that would mask the singer’s voice in performance. By pre-analyzing the passage and correcting the orchestration so that the strong expression remains while the audibility of the singer is maintained, I might be led to use an orchestration that was not my first choice. This new choice can evolve into a new musical character or a new way of thinking in the context of the whole piece.

The Score-Tool App therefore also serves as a tool to help me to find my artistic identity. With the tools currently available to composers, I found it hard to obtain the information I wanted to test the audibility properties of my scores. I needed new technical tools to help me with my artistic choices. I also assume that other composers are not fundamentally different from me and that some of my colleagues may share my experiences. Therefore, the Score-Tool may be beneficial to them as well.

OTHER OBJECTIVES

In general, the Score-Tool App makes it possible to explore new ways to analyze orchestration. The source code of my App is freely available, and with basic programming skills an additional component or algorithm can be added into the system. The Score-Tool App is platform independent, so the orchestration analysis is possible regardless of the system used. Nor does the use of the App require any commercial software or commercial library.

Possibilities for expanding the Score-Tool App are discussed in chapter “Future Research”. Currently, only a few systematic orchestration analysis programs are available, and an open-source app like Score-Tool has the potential to attract a community of musician-programmers to develop its features further.

THE BODY OF THE SCORE-TOOL PROJECT

The body of my project consists of three components.

1. This report describing the methods and background of the App and the artistic output
2. The Score-Tool App and its source code
3. An artistic output, the opera *All the Truths We Cannot See*, where the functionality of the App is tested.

The artistic output reflects the impact of this new knowledge on my own work as a composer. In the opera *All the Truths We Cannot See*, I used the Score-Tool App to help in my orchestrating so that the desired instrument or voice would be in balance with the orchestration. The analytical information also guides my artistic decisions when I compose. I describe the compositional process of the opera later in this report, as well as discuss how the use of the Score-Tool App has changed the way I work.

Part I BACKGROUND

1 ORCHESTRATION

Orchestration is a major part of the work of orchestral composition. It involves which instruments to choose, what instrumental combinations to use, the register of the instruments and voices, their dynamics, and even the seating arrangement of the players. Furthermore, the possibilities for writing even a single chord for orchestra are endless. In composing, decisions must be made for every piece, every section, every passage, every bar. In a way, the result of orchestration resembles the existing orchestral repertoire because that is the safe way to work; the repertoire is already known by orchestral musicians and audiences.

The fascination of orchestral music, whereby multiple instruments are playing individual parts concurrently, comes from the fact that the resulting "information overdose" is often challenging to our perception, a matter discussed later in chapter "For my project and for the Score-Tool App, I have included the possibility of testing the effect (or lack thereof) of seating on orchestration. The App includes a model of the main concert hall in Helsinki's Music Centre based on measurements made by Tapio Lokki and Jukka Pätynen. The measurement data are used with Lokki's permission. The effect of the listening position in the hall can also be tested in the App's tutorial section, where the user can move the individual instruments and the listener in a 2D space and see the effect of the position on the sound power in the instruments. In masking calculations, the average hall reverberation is taken into account by a 0.1-second overlap of changing notes.

Previous research on auditory perception related to the project". This overwhelmingly rich musical type is still our preference in Western orchestral music, because otherwise we would compose, play, and listen to something else. There are many aspects related to orchestration, starting with the acoustics of a hall, music psychology, music theory, psychoacoustics, a listener's experience, and so on. In this research, I decided to concentrate on my own experience as an orchestral composer and on how to analyze the perception of orchestral sound with a computer app.

2 METHODS

This is a multi-disciplinary research project in which I select the areas I want to include in my project and omit those I want to leave out. My decision has been to limit the Score-Tool App to the auditory perception of concurrent sounding orchestral chords and leave out, for example, the onset times of the orchestral chord pitches.

For this project, I borrow methods from different branches of psychoacoustics. My main expertise is in composition and writing music for orchestra. Along with drawing on my own experience as a composer as to how the orchestral apparatus works, I also provide an overview of some of the most well-known orchestration handbooks.

For the Score-Tool App, I use mathematical methods applied to sound to acquire numerical data about orchestration. The App also makes the data visible on screen so that users get feedback on their orchestration.

Here in Part I, I give a brief overview of the psychological aspects related to the audibility of the target, namely, the *cocktail party effect* and the *auditory stream* formation. These are subjects that play an important role in audibility, but they are not implemented into the current version of the App.

3 ORCHESTRATION HANDBOOK OVERVIEW

In this chapter, I discuss four orchestration handbooks selected either for their historical significance or because they are widely used today in academic settings to teach orchestration. My overview of these books is from the perspective of the Score-Tool project, with concentration on what the authors say about audibility, blending, and timbre in general. This overview provides a background to how academic composers generally learn. But composers also require considerable experience working directly with orchestras in rehearsals and performances, preferably of one's own works. The practical experience aspect is reflected in Part III where I discuss my new ideas and my artistic work in this Score-Tool project. Here, I begin by explaining why I chose these four handbooks for this overview chapter.

3.1 THE BOOKS SELECTED FOR OVERVIEW

Although hundreds of books and thousands of articles have been written about orchestration, there are four well-known opuses that are frequently quoted in articles and used as learning material in academic institutions. The first of these is Hector Berlioz's *Treatise on Instrumentation* (*Grand traité d'instrumentation et d'orchestration modernes*, 1844), enlarged and revised by Richard Strauss in 1904. This book has more historical than practical value, although many of Strauss's remarks in particular are still valid. The second handbook is Nikolai Rimsky-Korsakov's *Principles of Orchestration* (publ. 1913, written 1896–1908), a usable guide for a modern composer. The third is Walter Piston's *Orchestration* (1955), which contains analyses of various scores chosen by the author. And the fourth is Samuel Adler's *Study of Orchestration* (1982), with numerous exercises and an attached CD of examples in the latest edition. Adler's book goes from the basics to highly advanced orchestration techniques and has recommendations of scores for further study.

The one thing all these books have in common is that they do not provide ready-made solutions for how to orchestrate. Rather they encourage the reader to experiment and use inner hearing to develop the skills of an orchestrator. All four books have been organized the same way: they start by describing the author's relationship to the art of orchestration, continue with a presentation of the instruments in the symphony orchestra, and conclude by presenting some real-life examples from selected scores.

3.2 HOW AUDIBILITY IS ADDRESSED IN FOUR ORCHESTRATION HANDBOOKS

Next, I speculate on why the audibility of particular instruments is seldom addressed in orchestration handbooks. I also point out that the concepts of masking and blending are mixed in these books, and in some cases, it is difficult to tell which of the two concepts an author is writing about.

The audibility of a desired instrument seems to have been a secondary concern in orchestration treatises before World War II. This may be the result of certain conventions, i.e., having the first violins play a major part of the melodic material and using woodwinds to enhance the timbre while saving the brass for climaxes. In pre-World War II orchestration handbooks, remarks concerning instrument audibility are general, for example, describing a certain register of an instrument as "penetrating"[3] or cautioning against overpowering a weak-sounding instrument.[4] After World War II, there seems to have been a change of thought towards more

---

[3] Berlioz and Strauss 1948, p. 55.

[4] Berlioz and Strauss 1948, p. 74.

democratic instrumentation, because audibility issues begin to receive some attention in the handbooks. This can be seen in Adler's *Study*, in which he devotes a whole chapter to instructions for dividing the orchestration into fore-, middle- and background layers. This layered thinking includes the audibility aspect because Adler considers the foreground as the layer that is most clearly heard.

Notably, both Piston and Adler mention the overtone structure of sounds several times and point out how this structure can interfere with the rest of the instrumentation in different registers. This is the key concept of masking, but neither author goes deeper into the subject. Adler, for example, favors simply keeping the volume of the background layer down and suggests writing different musical textures for different layers.

With all four authors, audibility issues seem to mix with the concept of *blending*, which is often a desired phenomenon among composers. For example, the low-register flute is by turns presented positively as blending well with other instruments and negatively as being overpowered. The same applies to low-register strings, which the authors warn can be overshadowed by the brass, yet at the same time claiming that the brass do not blend with low-register strings.

The remarks about the effects of individual musicians diminish with time. While Berlioz and Strauss still point out the differences between weak and strong players in a specific passage, Piston in his examples treats the orchestra almost like a machine that always produces the same result with a given score. Strauss even stretches his concept to blending, claiming that "German oboe players produce a thick tone, which does not blend in with the flutes and oboes."[5]

Blending is more thoroughly discussed below (Part I, Chapter 7). Regarding the authors of the four orchestration books selected here, my impression is that, for them, blending means that sounds from the blended instrument cannot be recognized. An example of such blending would be adding a register to a pipe organ sound where the added sound, although coming from a different pipe, blends completely with the original sound, i.e., it cannot be recognized as an individual sound source.

Next, I take a closer look at the four orchestration handbooks and point out sections and remarks that have special significance for my Score-Tool project.

3.3 OVERVIEW OF THE BERLIOZ-STRAUSS AND RIMSKY-KORSAKOV ORCHESTRATION HANDBOOKS

The development of the romantic orchestra created a need for a comprehensive guide for composers regarding the full potential of such an instrumental group. One of the first such guides was Hector Berlioz's *Treatise on Instrumentation*. Berlioz addresses his fellow composers directly, wanting to share information about this wonderful apparatus, the orchestra. At the time of its publication (serialized in the early 1840s, published in full in 1844), the modern symphony orchestra was a relatively new concept, and Berlioz devoted much of his text to defining carefully the range and role of each orchestral instrument.

---

[5] Berlioz and Strauss 1948, p. 183.

His *Treatise* is not so much about orchestration as it is about individual instruments. The chapters are named after each orchestral instrument in use in the mid-1800s. Even rarities, such as the Double-Bass Ophicleide and the Saxotromba get their own chapter.

All information about orchestration, i.e., choosing and combining the instruments to obtain a desired timbre, is embedded in chapters on individual instruments. Concluding the book is a four-page chapter about the orchestra as a whole and a historically interesting chapter about conducting. From a composer's point of view, this is where the actual orchestration begins, but sadly this is where the handbook ends.

The original publication was a huge effort for Berlioz, because this was the first time anyone had attempted to summarize the conventions in orchestral music. We can even say that Berlioz's pioneer work affected the way orchestral music was written thereafter and encouraged composers to write for larger ensembles than before. For a long time, Berlioz's treatise was considered the best handbook available. Yet already in its own century its relevance began to wane.

Help came from Richard Strauss, a visionary orchestrator, who at the *Treatise* publisher's request, revised and updated Berlioz's book in 1904. Strauss commented mostly on the development of the orchestral instruments and reinforced Berlioz's remarks on topics still current in Strauss's time. The biggest revision was made in the chapter about horns: Strauss removed almost the whole original chapter and replaced it with an essay on the valve-horn, considering it obsolete given the rise of the valve-horn. Strauss's big contribution was to bring the music examples up to date by including scores of Wagner and some of his own.

A second significant orchestration book appeared just before World War I: Rimsky-Korsakov's *Principles of Orchestration*. This volume is written more freely than Berlioz's treatise. The first edition was released in 1913 in Russian, nearly seventy years after Berlioz's pioneer work, but just nine years after Strauss's revised edition of Berlioz. Rimsky's personal style of orchestration style is lush and colorful, making it an ideal subject for studying audibility and sound color. Already in the introduction, the author mentions the problem of detaching a melody from its harmonic setting.

Unlike Berlioz, Rimsky-Korsakov rushes through a presentation of the orchestral instruments to more interesting subjects, such as scoring a melody, harmony, and discussion of orchestral composition in general. Yet like Berlioz, Rimsky has a separate section for vocal writing. The vocal section is quite extensive, accounting for nearly one-fourth of the total number of pages and indicating the author's general interest in opera.

Rimsky's chapters about individual instruments follow Berlioz's example in giving a general overview of the range and role of each instrumental group. What stands out in that section is Rimsky's colorful attempt to characterize timbre. For example, in its low register the flute is described as "dull," the oboe "wild," the clarinet "threatening," and the bassoon "sinister."[6]. The characterizations are on par with Rimsky's composition titles, which are more descriptive in nature as opposed to titles for "absolute music."

Before Rimsky-Korsakov's *Principles* appeared, audibility and blending were not subjects on which Berlioz had concentrated in the world's first orchestration book. Yet there are glimpses of the subject here and there. The most frequent is a description of an instrumental sound as

---

[6] Rimsky-Korsakov 2013, p. 19.

"penetrating," indicating that it will be heard through the orchestration. The violin's e-string, the clarinet in a high register, the piccolo, and the trombone are all stamped with this description. This understanding is more or less intuitive now, but Berlioz appears to have been the first to express such information in written form.

The concept of blending is discussed by both Berlioz and Strauss, and not always in a positive manner. For example, when Berlioz says that the oboe's timbre is lost in the ensemble,[7] it is unclear if the "lost" means blended or inaudible. Either way it appears to be a negative effect. Elsewhere the expression is also ambiguous, as when Berlioz writes, "horns, trombones, and all brass blend best with the harp." Since the carrying power of the harp's sound is just a fraction of the power of any brass instrument, the statement can be understood either as the harp being overpowered by the brass or as sounding nice with them. I would have hoped for more comments by Strauss on blending in his annotations, but the only direct reference to the subject is in the horn chapter. There, according to him, the horn is probably the instrument that blends best with all instrumental groups.[8]

Among these authors, blending-related comments include observations on similarities and dissimilarities in sound colors. Berlioz was one of the first composers to play and experiment with orchestration techniques. His comments do not systematically address all orchestral instruments, but rather arise from his own practical solutions. Berlioz states, for example, that it is impossible to distinguish between cello harmonics and the violin e-string with a mute.[9] A bit surprising is Berlioz's remark that double basses combine very poorly with cellos two octaves above,[10] because cellos one octave apart from basses are known to blend extremely well. From Strauss, the surprise comes in his admission of an orchestration failure: doubling trombones with double basses is, according to Strauss, not only a poor blend, but also softens the trombone's effect.[11]

In his brief chapter on combining instruments, Berlioz warns about having too small a string section vis-à-vis the normal brass section with 4 horns, 3 trombones and 2 trumpets; against those, the violins would scarcely be audible. In the same chapter, Strauss claims that when a powerful sound is needed, one should consider using a combination of trumpets, strings, and woodwinds rather than solely a brass choir, because brass instruments balance each other by creating a softer sound rather than a mixture of different sound colors.

Rimsky-Korsakov discusses actual orchestration technique more than either Berlioz or Strauss. In Rimsky's *Principles*, there are very few examples in the instrument chapter about audibility, and one can only wonder if the examples come from the composer's own experiences. His advice, for example, is that a group of woodwinds with brass overpowers string pizzicato, piano, and celesta, but not glockenspiel, bells, or xylophone. This advice may apply to some special case, but it is clearly a matter of the register and dynamics used. The same criticism applies to Rimsky's claim that when a single woodwind is added to the string section, the woodwind sound will be lost.

---

[7] Berlioz and Strauss 1948, p. 164.

[8] Berlioz and Strauss 1948, p. 160.

[9] Berlioz and Strauss 1948, p. 78.

[10] Berlioz and Strauss 1948, p. 80.

[11] Berlioz and Strauss 1948, p. 97.

In general, Rimsky-Korsakov's approach to instrumental combinations is not very systematic, but arises from his own experience in exploring orchestration possibilities. He takes his examples mostly from his own compositions, largely drawn from the operas *Sadko* and *Snegourotchka* (*The Snow Queen*). These two works may not be the composer's most popular, but they seem to be significant to him, at least in terms of orchestral writing. The interesting thing is that the observations on audibility and timbre seem to come as a surprise to the composer himself, or at least Rimsky writes as if they do, wondering in his *Principles* why a passage sounds a certain way. This underlines the fact that orchestration is a complex web of dependencies, impossible for even the greatest names in orchestration history to master fully.

Rimsky includes some examples of timbre similarity. He states, for example, that the timbre of the viola in its middle register equals that of the bassoon or the clarinet in the low register, and that at the dynamic level *p* or *mf*, the bassoon and horn are "somewhat analogous."[12] He, however, makes a distinction between timbre similarity and blending, since a good blending of the oboe with the stopped horn or muted trumpet is expressed without referring to timbre likeness. Another example of this is when Rimsky first talks about the dissimilarity between string and brass tones, but then right away, points out that horns and cellos produce a "beautifully blended, soft quality tone."[13] For Rimsky, blending may thus be more of an artistic choice rather than making use of timbre similarity.

In his chapter on orchestrating a melody, Rimsky gives a few explicit examples of audibility. It is also interesting that these examples do not specify the dynamics, indicating that the effect applies to the whole dynamic range of the instrument. For example, according to the author, when the flute, oboe, and clarinet are playing in unison, the flute predominates in the low register, the oboe in the middle register, and the clarinet in the high compass.[14]

A unique approach by this author was to think of audibility as opposed to a tone's role in a harmonic context. For example, Rimsky urges the orchestrator to observe the oboe's penetrating tone when writing dissonances and gives notated examples of woodwind chords in which some instruments are too piercing, too prominent, or too weak. As with Berlioz, most of Rimsky's timbre examples involve woodwinds, and both authors openly admit that a woodwind group possesses the most interesting timbres in the orchestra.

Rimsky also adds a personal touch when he writes about audibility issues of the voice with the orchestra. Like Berlioz, Rimsky warns about using too heavy an accompaniment with the voice, but adds that too simple an accompaniment will lack interest and will not sustain the voice sufficiently.[15] Rimsky also lists in increasing order the orchestral instruments that are most likely to overpower the voice: strings, woodwinds, horns, trombones, and trumpets. The singer's voice would also drown in a timpani tremolo and in doubling any woodwind with the horn. Rimsky's advice is to avoid accompanying a singer with a specific combination of 2 clarinets, 2 oboes, and 2 horns. All this again suggests that Rimsky was listing his own past experiments with doublings.

---

[12] Rimsky-Korsakov 2013, p. 34.

[13] Rimsky-Korsakov 2013, p. 61.

[14] Rimsky-Korsakov 2013, p. 48.

[15] Rimsky-Korsakov 2013, p. 119.

Both Berlioz and Strauss composed a great deal of vocal music, but in the *Treatise on Instrumentation*, in both the original and Strauss's enlarged and annotated edition, there is only a brief chapter about vocal writing. The chapter deals with both choir and soloists and gives valuable insight into the balance problems Berlioz faced in composing for voices. In general, their advice is not to write parts in a low or even a middle register. There are also warnings about distracting the listener's attention with complex orchestration while voices are singing, and they advise avoiding doubling a vocal part with an instrument. Berlioz even criticized Bach's vocal writing in polyphonic passages accompanied with orchestra, where some parts cannot be "heard with clarity by the listener."[16]

Rimsky's vocal writing advice is more technical than Berlioz's. Even the singers' formant phenomenon is lurking in Rimsky's observation that "women's voices suffer more than men's when they come in contact with harmony in a register similar to their own."[17] The development of the carrying power of singers' voices from Berlioz's time is evident, since according to Berlioz, the clarity of the voice hardly tolerates any orchestration at all. Rimsky, on other hand, allows orchestration even in the same register as the singer.

Rimsky does not end his *Principles* with any conclusion or manifesto regarding the art of orchestration. Hence, I consider his treatise a written exploration of the subject rather than a wise man's words to the younger generation.

3.4 OVERVIEW OF PISTON'S AND ADLER'S ORCHESTRATION HANDBOOKS

Walter Piston's *Orchestration* from 1955 is, in its approach to the subject, almost the opposite of Rimsky-Korsakov's treatise. Whereas in almost hedonistic style Rimsky wondered about the possibilities of this "cornucopia of sound colors," Piston takes on the role of a schoolmaster to tell "how things are." Piston published a similar study on harmony and counterpoint, which suggests that he saw it as his mission to educate the younger, post-war era composers so that the tacit knowledge of his generation would not be lost. Piston apparently aimed for objectivity and carefully avoided talking about his own compositions or orchestrations. Piston also presented problems for the reader to solve and provided assignments to enable students to develop their orchestration skills. Both the target group and the subject Piston addresses is, in his words, the "student."

The structure of his book follows that of its predecessors, starting with the presentation of orchestral instruments and then moving to orchestration in general. The main difference from Berlioz's and Rimsky's handbooks is that Piston leaves out human voices altogether and concentrates on the basic selection of instruments in the modern symphony orchestra. In the instrument chapter, Piston clarifies not only the range and notation, but also the physical sound-producing mechanism of each instrument.

In the middle of his handbook, Piston makes perhaps the best remark about orchestration that I have ever seen:

---

[16] Berlioz and Strauss 1948, p. 353.

[17] Rimsky-Korsakov 2013, p. 121.

> Well-known and well-loved symphonic masterpieces are made to sound well orchestrated only through sympathetic and understanding adjustment of the written parts in performance.[18]

In other words, orchestration written down on a sheet of paper comes alive only through a careful and respectful performance. This view is repeated later, when the author, discussing Debussy's *L'après-midi d'un faune*, observes that the accompaniment could overwhelm the melody unless performed with understanding.[19] These wise words go counter to what is often done today: read-through performances of scores of young composers in which the music is played prima vista, without a proper rehearsal period. Without the sought-after sympathetic adjustment of the written parts, a read-through may cause more harm than good to the orchestrator. This is one reason I developed the Score-Tool App: to make it possible to obtain an objective, computer-generated feedback of the orchestration by approximating the performance indicated in the score. Later, Piston even added that the conductor who brings out voices intended to remain in the background is an all-too-familiar figure.[20] This comment may refer to conductors who emphasize their own interpretations by making the piece sound "unlike any other performance before."

After Piston, composers developed their orchestration techniques by experimenting with timbres and textures. These techniques rendered Piston's book a bit old-fashioned. Help came from Adler's *The Study of Orchestration*, which is the most comprehensive of all the four handbooks examined here. Its first publication came in 1982, but Adler constantly updates his opus: currently, the fourth edition of the book was released in 2016. Its structure follows the well-known formula: first presenting individual instruments followed by a general discussion of orchestration with score examples. However, Adler goes further in his chapters than his predecessors, aiming to create both a study book for the student and a reference book for the professional. Like Piston, Adler does not present voices and their abilities in the beginning, but instead includes a later chapter on accompanying a vocal soloist or chorus. In addition to the most common orchestral instruments, Adler deals with a few rarities, such as the musical saw and the cimbalom.

The presentation of the instruments is a good review for any composer, a chance to remind themselves of all the technical possibilities one can require of each player. Adler discusses even experimental playing techniques, such as quarter tones and blowing air through reed instruments without a reed. He also adds to each instrument chapter a few excerpts from familiar scores about how to write idiomatically for the instrument.

After World War II, many composers abandoned the romantic idiom, at least for some years, and began giving their works neutral and scientific titles rather than poetic ones. This trend can already be seen in Piston's book, where, in a few instances, the author tries to take an engineer's role in describing the sound but fails to provide adequate statements. One example is his claim that the lost power of second violins, when seated behind the firsts, is one of many acoustical problems standing in the way of an exact science of orchestration.[21] The fact is that, in a hall, seating order is irrelevant in terms of audibility if the distance to the listener is, say, ten meters

---

[18] Piston 1955, p. 121.

[19] Piston 1955, p. 365.

[20] Piston 1955, p. 374.

[21] Piston 1955, p. 61.

or even more.[22]Another misunderstanding appears in Piston's claim "An acoustic phenomenon to be noted when the solo violin plays in the midst of a large orchestra is its surprising ability to make itself heard."[23] If the phenomenon is acoustic, it might refer to certain example where the spectrum created by violin sound do not overlap with the spectrum of orchestra. If it is said in general the effect might come from the field of psychology.[24]

The concept of blend does occur occasionally in Piston's text. We get a hint of the author's definition of the word when he states that the sounds of the English horn and solo cello blend as one, and he goes on to say that neither predominates at any time. This suggests that Piston thinks of blend not only as timbre, but also as an issue of loudness. In one place, it is unclear whether Piston is referring to blending when he claims that in d'Indy's *Istar* "Four horns, playing legato, are absorbed in the over-all sound."[25] Absorption described here seems like blending.

Piston goes further than the two previous authors in addressing audibility issues. For example, with the violin he makes a distinction regarding which strings are used. According to him, the e-string has the most carrying power, while the least powerful is the d-string. The timbre properties of each string can also be explored in the Score-Tool App, which includes analysis data for each string played separately.

Piston's observations can also be contradictory, as when he first describes the top string of the viola as nasal with a tendency to sound unduly prominent, yet a few pages later states that the viola's tone quality is such that it is easily covered by accompanying sounds. In another place, he describes the muted trumpet as the most piercing sound in the orchestra, while later he says that the muted trumpet blends especially well with the English horn. In the latter example, it of course depends on dynamics.

Piston's partly scientific approach can be observed in the sentence in which he points out that strong overtones from low bass notes can cover the flute's first half octave.[26] This phenomenon is a result of auditory masking, the very effect that is the focus of my project. In my view, auditory masking creates the most audibility surprises that a composer faces in orchestration, and I am pleased that the subject was acknowledged already in the 1950s. Auditory masking also plays a role in Piston's observation that the bass drum and triangle can be heard even in the loudest *tutti*. In his book Piston goes no further than that, but the reason for the audibility of these percussion instruments lies in the fact that they stimulate the very low and very high auditory bands where other orchestral instruments' sounds seldom reside.

Piston's most interesting comments regarding audibility include a remark about a passage in Debussy's *La mer*. Piston wonders about the flutes, which gauging by the orchestration, seem unlikely to be heard; yet performances prove that the flutes sound here with extraordinary clarity.[27] In my opinion, this is again proof that orchestration is a complex matter, even for

---

[22] See Part I, Chapter 10.

[23] Piston 1955, p. 63.

[24] See Part I, section 11.2.

[25] Piston 1955, p. 358.

[26] Piston 1955, p. 131.

[27] Piston 1955, p. 373.

experienced professionals. Whether Debussy knew that the flutes could be heard in this passage remains a mystery. Another of Piston's notable claims is that in *tutti* chords, brass instruments render woodwinds useless in the middle register.[28] In most cases when an instrumental sound is masked, it can still add components to the overall harmony, even though the instrument's actual sound is not heard. The claim that there are cases where scoring for a certain instrument is useless is a strong one and is not further mentioned by the author, also he does not distinguish masking and blend. If the sound is removed, and an audible change can be detected, then it's blend not masking.

The final example I will mention from Piston's treatise is almost a punch in the face for a careful orchestrator: according to Piston, in *tutti* chords the temptation to mark different dynamic levels for brass and woodwinds must be resisted.[29] This comes back to the importance of sympathetic adjustment, i.e., performing music with the context in mind. Yet as a composer I would like to have as much control as possible over the balance of a chord. The comment must perhaps be weighed against the fact that the target audience for Piston's book is the student composer at the beginning of a career.

In Adler's *Study of Orchestration*, the technical remarks are well-structured. There are almost no audibility-related remarks in the first part of the book except those concerning the harp, about which Adler makes a thoughtful statement. The author has realized that the weak-sounding harp has more sound potential when the pedals are in the flat position, because then the string is the longest. However, in the *Study*'s second part, audibility is discussed more than Berlioz–Strauss, Rimsky-Korsakov, and Piston combined. Adler even advises dividing the orchestral score into three categories: fore, middle, and background. For Adler, this is perhaps mere thought-play because there are no explicit instructions about how to orchestrate background or middle ground. To clarify his point, Adler uses specific examples; for example, in discussing Brahms's *Symphony No. 3*, he mentions that "Many professional orchestrators recommend staying away from the register of the melody line. In many instances, especially when foreground and background instruments are played by instruments of similar color, this is good advice."[30] This is again the key point of auditory masking, although Adler speaks about the "fuzzy concept of sound color" when, in my view, the better choice for that term would have been frequency structure.[31]

Although Adler is clearly aware of how auditory masking works, he still, perhaps for reasons of simplifying the expression, advises softening the dynamics of all the other instruments whenever a weak-sounding instrument is playing. An example is when he warns about the thin sound of a bassoon in its high register: "The upper fifth of the register does not project well; the dynamics of the accompanying instruments should be soft enough not to overshadow the soloist."[32]

---

[28] Piston 1955, p. 449.

[29] Piston 1955, p. 445.

[30] Adler 2016, p. 123.

[31] The both concepts are intimately related, but I personally try to avoid linking sounds and colors together, because it would be misleading to, for example, talk about red or yellow sounds.

[32] Adler 2016, p. 222.

Like Piston, Adler addresses the importance of balancing the score in performance. He points out that in a passage with the familiar combination of clarinet and oboe in unison, the conductor will have to balance the two. Luckily, unlike Piston, Adler does not advise avoiding different dynamics for different instruments. This may be a result of changes in performance habits after World War II. Before that, mixed dynamics in *tutti* chords would perhaps were considered a mistake. [33]

Regarding doublings with the same instruments, Adler makes a good point when he questions whether in a *tutti* passage a single instrument alone would be strong enough to carry the melody. In what instances would we double the melody with another player of the same instrument? In other words, when do two flutes sound louder than one? Experts disagree on the exact answer.[34] It is well-known that two flutes playing the same part will not double the volume as compared to one player but will merely raise the volume approximately 3 dB. Depending on wall reflections and the player's position, the sounds may even start to cancel each other out. Adler's main point seems to be that it is seldom necessary to have two players on the same part. Three to 24 players on one part is totally different but having two on a part is generally a bad idea, especially in solos, because two players can distract from each other's expressiveness.

With regard to horns, other authors have pointed out the ability of horns to blend in. This view is endorsed by Adler, but in addition he claims that horns playing in unison with woodwinds strengthen the woodwind sound. This idea of blending differs a bit from Piston's idea. Piston seems to view blending as an equal merging of two sounds, but based on this example, Adler apparently thinks that, after blending, the woodwind sound remains the most prominent sound color. Adler's concept is what Sandell calls for augmented timbre.[35] Adler's point that the piano mixes relatively well with all orchestral instruments but blends with none is a bit mysterious because he does not clarify what "mixing well" means. Mixing could in some cases be analogous to blending, but Adler uses it here as the opposite of blending. Adler stirs the soup even more when, near the end of his *Study*, he states that some instruments tend to blend in with the orchestra when played in certain registers and with acoustically sympathetic orchestral combinations.[36] "Acoustically sympathetic" is an ambiguous term and could mean either matching or contrasting frequency content. It emerges in the text that this is an unwanted phenomenon, since the author advises resisting this tendency.

In many instances, Adler stresses the importance of a distinct timbre in order to set off a desired instrument, i.e., with a timbre of the target instrument that differs from the surrounding orchestration. His examples include Tchaikovsky's *Francesca da Rimini*, mm. 325–349, where the color of the accompaniment (a pizzicato string orchestra) contrasts with that of the solo instrument (a clarinet), and Wagner's *Meistersinger Prelude*, where the three elements (three different melodies in counterpoint) create a well-defined passage with each assigned a distinctive color combination.[37] By three elements, Adler probably means three melody lines orchestrated for different instrumental groups in the work. What I understand from his

---

[33] See Part II, section 5.3.

[34] Adler 2016, p. 238.

[35] Sandell 1991, p. 32.

[36] Adler 2016, p. 635.

[37] Adler 2016, p. 584.

examples is that Adler means some kind of negative blending, i.e., the sounds have such distinctive timbres that they are perceived as separate entities. This concept is called *stream segregation*. It is an interesting and seldom discussed effect in orchestration that would probably need a listening test to determine the amount of timbre difference needed for the phenomenon to happen.

Adler's remarks on audibility resonate with other authors, as he stresses the power of brass and high piccolo while warning about low flute and harp. He gives several insightful examples from grand masters, where the orchestration does not work as intended. One is from Schubert's Eighth Symphony, in the first movement, where Adler points out that in m. 26 the flute goes almost undetected, not really heard until it plays the highest notes of the cadential chords.[38] He also relates a humorous anecdote once told by Piston to his students: when you write a *fortissimo* for the timpani doubled by the bass drum, don't expect to hear anything else from the rest of the orchestra. The statement is perhaps exaggerated, but the truth is that masking spreads heavily upwards in frequency space, resulting in a loud bass instrument occupying several auditory bands.[39]

Adler's discussion of one example taken from the first movement of Brahms's Third Symphony expresses a view about which I have some doubts. In the final *tutti* chord, according to Adler, the flutes can easily overpower the chordal structure. All dynamics are marked *p*, and the flutes are in their optimal register. It has been my experience that, unless the flutes force the dynamics immensely louder than indicated, the flute sound will not likely overpower the chordal structure. Everyone can easily make their own decision by listening to different interpretations of the chord.

In one of the last chapters in his *Study*, Adler discusses the role of singing voices in orchestration. The examples include both solo voices and choir with orchestra. Whereas Berlioz and Rimsky-Korsakov were afraid of adding nearly any orchestration to the singer's passages and Piston advised orchestrating lightly, Adler seems to think that the voice will sound even through a thick texture. This becomes apparent in the section in which Adler is more concerned about the beauty of the voice than with actual audibility: "Forcing the voice to sing over a thick orchestral accompaniment for long periods of time overtaxes and fatigues singers besides causing them to strain abnormally, all of which detracts from the beauty of the voice's natural quality."[40] In my experience, this kind of orchestration with voices has been evident in many contemporary operas, including my own, where the need for orchestral power outweighs sympathy for the singer. This is one more problem where the Score-Tool App c an provide at least a good guide to a solution.

3.5 CONCLUSION OF ORCHESTRATION HANDBOOK OVERVIEW

Combining the knowledge found in orchestration treatises from different eras, my conclusion about target audibility is the following:

- An instrument's sound is considered audible when the instrument's characteristic timbre can be identified from the mass, either as non-masked or as a part of an augmented timbre.

---

[38] Adler 2016, p. 234.

[39] See Part I, section 5.5.

[40] Adler 2016, p. 639.

- Sound is blended when an instrument's sound is not clearly recognized, but it is a part of an emergent timbre.

As stated in the beginning of this chapter, there are numerous books written about orchestration, including from very technical points of view and with focus on the spectral qualities of instruments[41]. However, none of those books concentrates specifically on my research field, masking and blending. Therefore, I decided not to include these handbooks in this report, but instead to examine the latest research on orchestration. But first, I briefly review the basic acoustic and psychoacoustic concepts essential for understanding the Score-Tool project.

4 ASPECTS OF OUR HEARING SYSTEM RELATED TO THE PROJECT

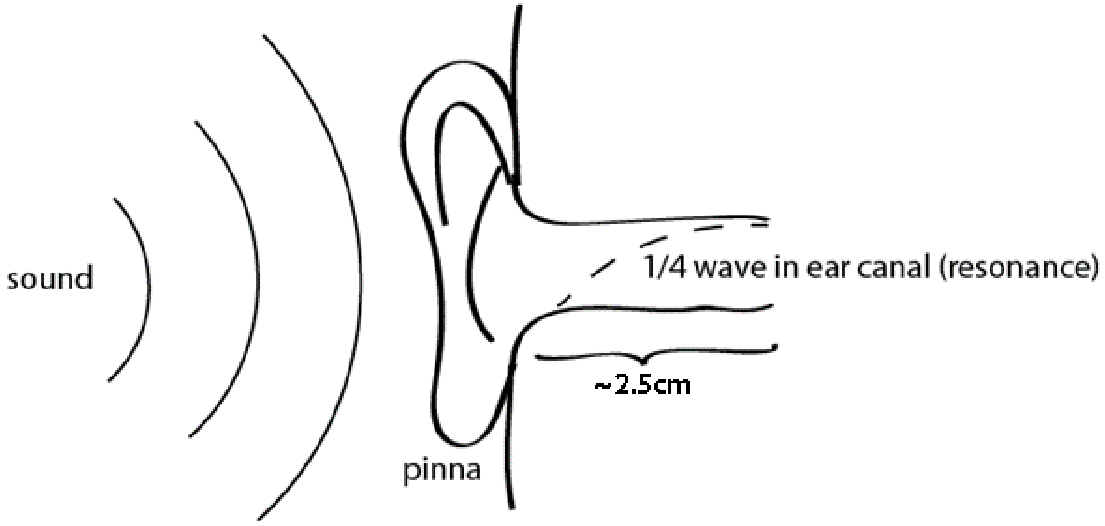
In this chapter, I present the basic functions of our hearing system, including partly in detail, because in my view it is relevant that the algorithms I use have a firm basis in what happens inside the ear when a sound wave enters. A reader not interested in this subject may very well skip this chapter and still obtain a good view of the Score-Tool project.

The human hearing system can be divided into two regions very different in character. The first region starts when a sound enters the ear, and the oscillations are pre-processed in the outer, middle, and inner ear. The second region starts beyond the inner ear, where the oscillations are encoded into electricity.[42] In the second region, the electric potentials are interpreted by neurons, which ultimately lead to auditory sensations. What happens in the second region is partly still a mystery, and research on that is an ongoing process. For practical reasons, in my project I concentrate on pre-processing models in the first region, where mathematical formulas can be applied to an oscillating wave. Because of unknown factors of processing in the second region, it is not possible to give general answers about target audibility in orchestration.

---

[41] Spectral qualities of instruments are discussed in detail in Part I, chapter 5, "Concepts used in the report"

[42] Fastl and Zwicker 2007, p. 23.

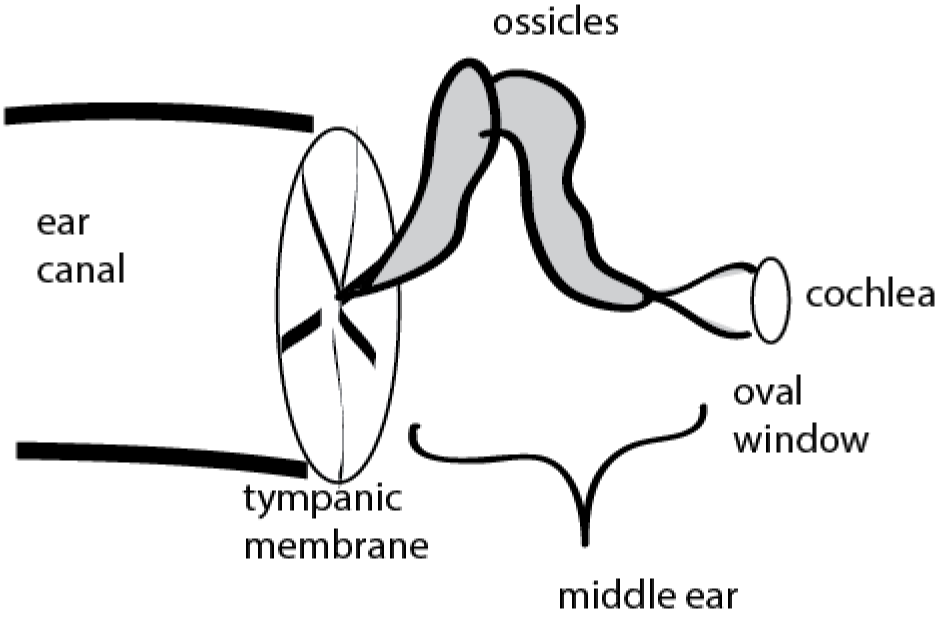


Example 6. The pinna receives the sound direction, and the ear canal resonates especially on a quarter of the wavelength in the 3.5 kHz frequency range

The first region of our hearing system consists of three sections. The sound pre-processing begins in the outer ear, where the sound reflections caused by the *pinna* – the broad flap of skin-covered cartilage which forms the external ear – , give the first clues to the direction of the sound. When the sound wave enters the ear canal, the canal acts as an acoustic resonator, and it has a strong influence on the frequency response of the hearing system. The length of the canal, about 2.5 centimeters, corresponds to a quarter of the wavelength of 3.5 kilohertz (hereafter kHz) frequency ($\frac{\text{sound speed 343 m/s}}{\text{3500 Hz}} = 9.8cm$).[43] This results in high sensitivity in this frequency range, which, when converted to a musical note, corresponds to approximately the highest keys on a full piano keyboard. This frequency range is so sensitive that the loud sounds in that particular range are easily experienced as disturbing, which is the reason that fire alarms and other alerts sound very loud, namely, in the 3.5 kHz frequency area. The loudest frequency peaks of orchestral music are situated in a much lower region, which largely makes it possible for us to listen to music for longer periods of time. However, in music, the 3.5 kHz area is important for the sensation of timbre or so-called “sound color.”

---

[43] Fastl and Zwicker 2007, p. 24.



Example 7. The middle ear's ossicles match the sound wave impedance between air (ear canal) and liquid (cochlea). The sound transmission is most effective at frequencies from 500 Hz to 5 kHz.

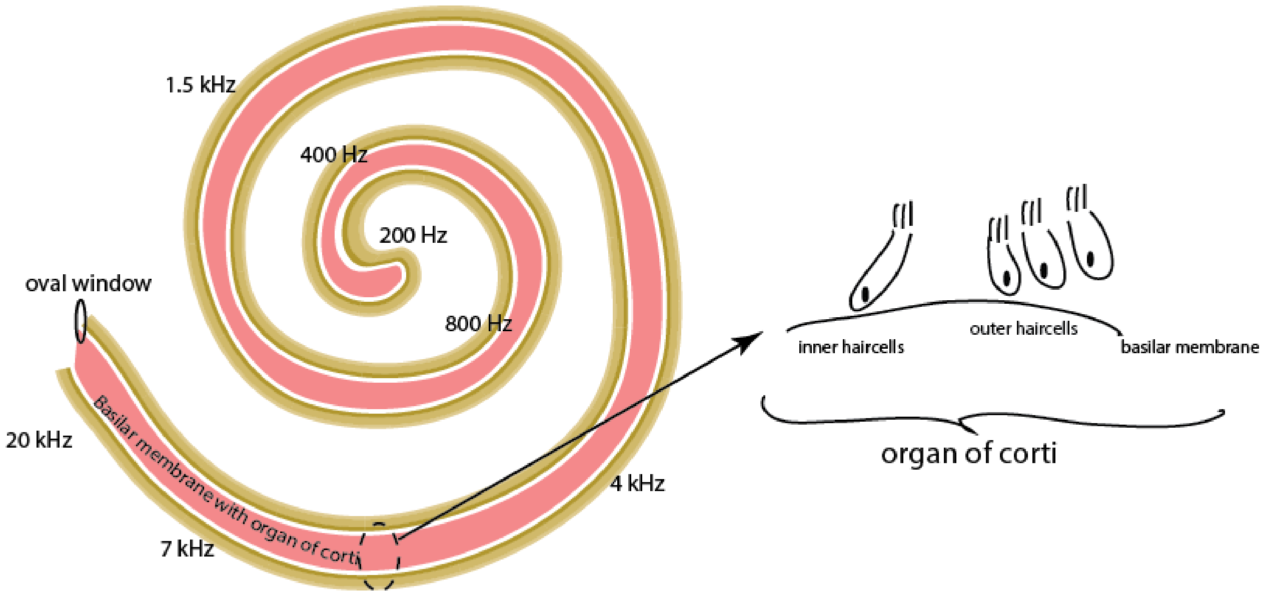
In the middle ear, there is a bone structure responsible for transmitting the sound waves from air in the outer ear to the liquid in the inner ear. This transmission happens in different stages, but a closer look at the stages is unnecessary for the Score-Tool project. The important part is that, on entering the liquid, there is an impedance change in the sound waves. Technically, the middle ear acts as an impedance-matching device. Transmission of sound through the middle ear is most efficient at middle frequencies (500-5000 Hz).[44] This phenomenon makes an additional boost in hearing sensitivity around the already resonant 3.5 kHz frequency area.

The most important part for hearing in the inner ear is the *cochlea*, which is shaped like a snail. Inside the cochlea the sound travels through liquid to the *basilar membrane*. The cochlea forms 2.5 turns, allowing a basilar membrane length of about 32 millimeters (hereafter mm) On the basilar membrane is the *organ of Corti*, where tiny cells, called *hair cells*, detect different frequencies and encode them into electricity for the brain to interpret. There are two kinds of hair cells, inner and outer. The outer hair cells refine the area of stimulation on the basilar membrane, which enhances the detection of frequencies. The inner hair cells react to movement of the part of basilar membrane, which leads to detecting the frequencies. There are approximately 3,500 inner hair cells and 12,000 outer hair cells uniformly distributed along the organ of Corti.[45]

---

[44] Moore 2012, p. 24.

[45] Pulkki and Karjalainen 2015, p. 117.



Example 8. The inner ear's cochlea. Sound enters the cochlea via the oval window. Within the cochlea is the basilar membrane on which is located the organ of Corti. This organ includes hair cells that detect and amplify sound waves when the basilar membrane resonates. Different frequencies are detected on different parts of the cochlea.

Sound waves stimulate the cochlear fluid and travel along the basilar membrane as a traveling wave. The curve that shows the amplitude of the traveling wave at each point along the basilar membrane is called its envelope. The base of the basilar membrane has a high degree of stiffness and low mass. By contrast, the apex end has a low level of stiffness and high mass. Because the stiffness and the arrangement of the hair cells affect the cells' response, the inner hair cells at the base end respond to high frequencies, whereas the inner hair cells at the apex end respond to low frequencies. The hair cells along the basilar membrane respond to the movement; therefore, the resonances of different frequencies at different points on the basilar membrane dedicate the hair cells to respond to particular frequencies.[46] Thus, the high frequencies are detected first, and the low frequencies are detected last, when the sound wave has already travelled through the whole basilar membrane in the cochlea.

Because discrimination between different frequencies is mechanical, detected by the inner hair cells from the resonant vibration of the basilar membrane, the resolution of hearing is not very high at this point in the hearing system. The frequency-dedicated inner hair cells respond very well to their best frequency, but also respond to some extent to the second best, third best, and so on. The strength of the response depends on the power of the sound wave; the greater its power, the greater the response. The frequency detection happens by temporal coding through the rhythm of firing of nerve impulses, which captures aspects of the signal as filtered on that part of the basilar membrane. The total image of what we hear is constructed in the nervous system by the activation pattern of the inner hair cells.

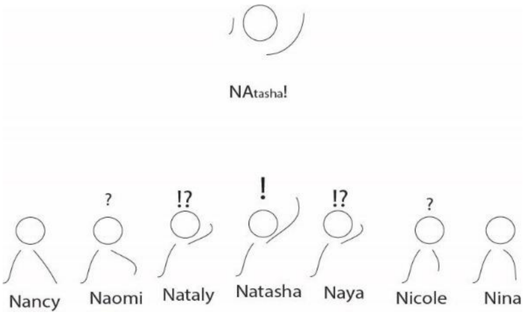
---

[46] Gelfand 2009, p. 87.

The detectable frequency range in humans is commonly said to be from 20 to 20,000 Hz. Although in laboratory conditions some people have detected frequencies in a range from 16 to 28,000 Hz,[47] the practical range is narrower. The components of music reside approximately in the range from 40 to 10,000 Hz,[48] which makes it possible to enjoy music even with some hearing loss in the high frequency area.

The just-noticeable variation in frequency is about 3 Hz up to 500 Hz; thereafter, it rises steeply to about 100 Hz at the 10 kHz point.[49] From this, it is apparent that the variation in frequency is not detected by just one hair cell, but by a combination of the activation of different inner hair cells. When one hair cell is activated, it is "busy" and cannot respond to further stimuli. It is thus impossible to hear two distinctly different sine waves that are close in frequency. Instead of two sounds, we hear roughness or beating. The beating of close frequencies creates an amplitude modulation pattern in that region of the basilar membrane, and the temporal firing rhythm captures the amplitude modulation pattern that is interpreted as roughness.

The hair cells that are busy responding to one frequency area at a time can be thought of as a line of people in alphabetical order by name. When a name is shouted, the cell with exactly that name responds immediately, while the cells with almost the same name respond uncertainly, as illustrated in Example 9. On the other hand, if the name is whispered into a person's ear, people nearby with almost the same name do not respond.



Example 9. Alphabetically-ordered people as an analogy for dedicated hair cells. When a name is shouted, several people respond.

When the basilar membrane envelope resulting from the spectrum of a more intense sound includes the envelope resulting from a less intense sound creates the effect of masking. A busy hair cell, even if it is busy with a "wrong" frequency activated by a certain amount of power, cannot detect another frequency with less power until it is free. This phenomenon means that the frequency discrimination range of the human hearing system consists of much larger units than fractions of frequencies or even single frequencies. The large blocks of inner hair cells responding to a certain frequency area are called the *auditory bands* or *critical bands*.

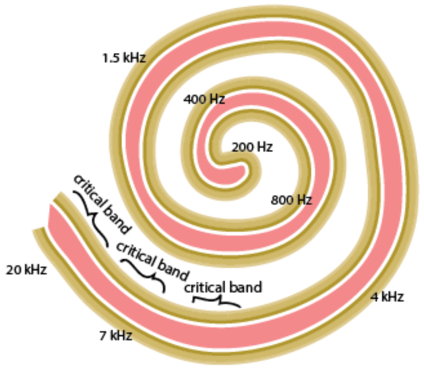
---

[47] Ashihara 2007 (published online)

[48] Fastl and Zwicker 2007, p. 18.

[49] Fastl and Zwicker 2007, p. 183.

The width of the critical band corresponds to the number of inner hair cells activated by the frequency. The areas where the basilar membrane resonates to a certain frequency are well-known, thanks to different research approaches. One approach is to place tiny microphones inside the active cochlea, and another is to conduct various listening tests with band-passed noise. Starting with the high frequencies, which respond to the sound first in the cochlea, the critical bandwidth is roughly 20% of the center frequency.



Example 10. Critical bands are the result of multiple inner hair cells fired by traveling wave on the basilar membrane inside the cochlea.

5 CONCEPTS USED IN THE REPORT

5.1 PITCH AND FREQUENCY

The sounds used in my program are all musical sounds with or without a pitch. Pitched instruments are notated according to the perceived pitch of the instrument's tone. The notation of non-pitched sounds is ambiguous, but if there are several non-pitched instruments in the orchestra, usually some kind of relative pitch notation is used.

If we consider the sound of a pitched instrument, we usually hear one prominent pitch, which corresponds to a certain number of cycles per second. The unit for 1 cycle per second is 1 Hertz (hereafter Hz); in other words, a periodic signal of a wave of 1 cycle per second has the frequency of 1 Hz.

The musical unit for frequency is a note, expressed in musical notation, or a letter with an octave marking. Unlike Hz, a note does not indicate an absolute frequency value, but needs a tuning reference, for example, middle c on the piano or C4. C4 has the rounded frequency of 261.63 Hz in equal temperament tuning, where A4 is tuned as 440 Hz. If the temperament or reference tuning of A4 were to be different, then the rounded frequency of C4 would also change. The Score-Tool App uses an equal temperament tuning of 440 Hz as A4, although this results in minor glitches in the sound representation of low instruments, because the intonation of individual players in the recording varies slightly (440-442 Hz A4), and errors are most prominent at low frequencies.

5.2 HARMONICS/OVERTONES

Pitched instrument sounds consist of multiple sinusoidal waveforms, which are usually in harmonic relationship to one another. "Harmonic relationship" means that the sinusoids are multiples of the base frequency. The multiples of the base frequency are called *harmonics*, *partials*, or *overtones*, and the whole complex is an instrument's *spectrum*. The base frequency is usually the pitch of the instrument's sound.

If the sinusoids of a sound are close but not in perfect harmonic relationship to one another, the sound is called inharmonic. Inharmonicity happens in all instruments that are plucked or played with mallets, because the strings and plates resonate in multiple dimensions. If the tension of the sounding body is loose, it will produce higher amounts of inharmonicity than a stiff body. The low keys of the piano, for example, produce more inharmonic partials than the high keys.[50]

Some orchestral sounds have a spectrum that is not even close to a harmonic one. Those instruments include, for example, drums and cymbals. This kind of sound is called noise, as its spectrum is completely unpredictable.

5.3 LOUDNESS

Sound waves produced by orchestral instruments are changes in pressure in the medium, which is air. The pressure of the sound can be measured in pascals (pa), which is the unit of pressure in any medium. However, a pascal is not a good unit for musical sound because of the logarithmic nature of the sound pressure reception in our hearing system. A better unit was invented in Bell Labs in the 1930s, a unit which is a logarithm of sound pressure in pascals, called Bel. Our hearing system can detect sounds from 0 to approximately 14 Bels. To avoid

---

[50] Järveläinen, Välimäki, and Karjalainen 1999.

the use of unnecessary decimals, usually a unit of a tenth of a Bel is used, called a deci-Bel or dB. If the dB value indicates the sound pressure level, then it must be indicated, for example, with an SPL (sound pressure level) marking. The complete formula for converting the pressure of pascals into deciBels is the following, including the reference sound pressure in air, which is 20 $micropascals$: $20\log_{10}\left(\frac{measured\ pressure}{reference\ pressure}\right) dB$

The detectability of the sound pressure is not constant through the frequency space of the hearing system. For example, a sound of 0 dB with a frequency of 4 kHz can easily be heard, but a 0 dB sound with a frequency of 100 Hz is completely undetectable. This non-linearity comes from the various physical conditions of the outer-, middle-, and inner-ear, and is somewhat similar in all hearing systems in which there are no disabilities.

The psychological term describing the sensation of pressure in the hearing system is called Loudness. Loudness levels correspond very well to the dynamic notations in music of *forte*, *piano*, *mezzo-forte*, and so on. The loudness of a sound depends on several factors such as rate of firing, which fibers are firing, the synchronization among the firing fibres, and on the total number of nerve impulses that reach the brain per second along the auditory tract.[51]

For the conversion from sound pressure to loudness, equal-loudness contours are used, which are obtained from a listening test conducted with a large group of subjects using variable sound stimuli. These contours are commonly used in acoustic applications, and they are proven to describe the sound pressure sensation well.[52] An example of a graph of these contours is given in Example 11.

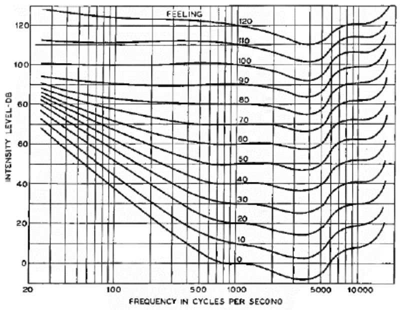


Fig. 4.—Loudness level contours.

Example 11. Equal-loudness contours from a hearing test in Bell Labs experiment in the 1930's. The contours are still valid.

5.4 CRITICAL BAND

As mentioned earlier, when the basilar membrane envelope resulting from the spectrum of a more intense sound includes the envelope resulting from a less intense sound. If the amplitude of the traveling wave is large, it activates a cluster of hair cells creating a spread. So, the amplitude affects the width of the spread. This is the direct consequence of the basilar member's resonance. Loud amplitude creates bigger resonance, which affects a wider area.

---

[51] Fletcher and Munson 1933.

[52] Fletcher and Munson 1933.

The number of critical bands in an individual hearing system is unique, but there are several well-known, good approximations that are used in acoustic applications. In my App, I use Bark bands, which divide the frequency range of our hearing system into 24 critical bands.[53] This model is a good representation of critical bands discounting borderline cases. A borderline case is when a frequency is situated close between bands.

In my Score-Tool App, I have used the MPEG psychoacoustic model, which divides the hearing range into 108 bands, the fractions of the Bark bands. The App keeps track of which sub-band belongs to which Bark band.[54] Thus, when using a 108-band model, one must keep in mind that one band does not equal a critical band, but only approximately one-fourth of it.

5.5 AUDITORY MASKING

Auditory masking is a phenomenon whereby a soft sound is inaudible in the presence of a louder sound.[55] Inaudibility of a weaker sound happens especially when two sounds have roughly similar spectral content, because they create similar kind of basilar member envelopes, of which less intense is masked. A situation in which a loud sound masks a soft sound with non-matching spectral content is called *spectral masking*. Spectral masking can occur even though the spectra of two sounds do not overlap. If the soft sound is still audible in the presence of a loud sound, but the perceived loudness of the soft sound is reduced, then *partial masking* occurs.

Auditory masking in orchestral music is mostly partial or spectral masking, but there may be cases of *informational masking*, which is discussed later. We might even assume that partial masking happens nearly every time that two or more orchestral instruments play simultaneously unless the instruments' spectra do not overlap at all.

5.6 SPECTRAL MASKING

Spectral masking thresholds have been widely tested with pure sine wave tones (test tones) and noise. White noise is a sound that contains all frequencies in an audible range in equal measure. If a sine wave tone cannot be heard in the noise background, the SPL level of the sine wave is then under the masked threshold. In the presence of white noise, the masked threshold of the test tone is constant at low frequencies up to about 500 Hz. Above that level, the threshold rises with increasing frequency. At 10 kHz the masked threshold is about 10 dB higher than at 500 Hz.[56] Although white noise has very little in common with orchestral sound, the 500 Hz turning point is important, since the average distribution of energy in the sound of orchestral music peaks at 300 Hz.[57] Thus the frequency region with a constant masked threshold, that is, the area under 500 Hz, is of great importance for orchestral music, although harmonic orchestral sounds consist of complex tones, not pure tones.

Listening tests conducted with narrow band noise show that the masking effect is strong when the masking noise and the masked tone are on the same critical band. The masking effect

---

[53] Smith and Abel 1999.

[54] ISO/IEC 1996.

[55] Pulkki and Karjalainen 2015, p. 156.

[56] Fastl and Zwicker 2007, p. 63.

[57] Sundberg 1977, p. 89.

spreads, however, to neighboring critical bands. This spreading can be modeled by a frequency slope, which is narrow at frequencies over 500 Hz; in other words, it spreads only a little and widens at frequencies under 500 Hz, meaning that it spreads more. The amount of spread also depends on the loudness of the masker, with louder masking noise generating a wider spread. [58]

Tests with noise masking a tone are popular because our ears are much more sensitive to tones than to noise. Noise is often considered a disturbing element, while tone is considered something that contains information. Orchestral music, and music in general, consists mainly of tones with pitches, with the noise element present mostly in percussion instruments and in extended techniques on string and wind instruments. Situations where an instrument sound is masking another instrument sound are therefore far from noise-masking-tone listening tests.

5.7 SINE WAVE MASKING SINE WAVE

Masking experiments of a sine wave tone masking another sine wave tone have proven to have many issues.[59] In the presence of a tone, any sensation which alters the perception of the tone has an effect. If the masker and the masked tones are at the same critical band, we hear beating or roughness when another tone is present. This happens even when another tone is masked and inaudible as a tone, although if beating is heard, the tone is not fully masked. This kind of "fusion" of two stimuli at the same critical band is called concurrent integration. In listening tests, the subject responds to this kind of situation, although the criterion is different than hearing an additional tone, i.e., when masking does not occur.[60]

Additionally, the difference tone produced by two adjacent tones disturbs the hearing experience. The difference tone is produced through nonlinear distortions that originate in our own hearing system.[61] In listening tests, the subject often hears the difference tone, even though the original tone is masked. Audible difference tones occur especially when the sounding tones are at frequencies between 1 and 2 kHz.

As with the narrow band masker, the tone masker also produces spread to adjacent critical bands. Interestingly, the frequency slope of the spread is greater towards the low frequencies at low loudness levels. At high levels, this behavior is reversed, so that a greater spread of masking is found towards higher frequencies. The effect at low levels is rather unexpected. At a 20 dB masker level, more spread of masking towards lower frequencies occurs. At 40 dB, masking is symmetrical, and at 60 dB there is more spread towards higher frequencies. Put another way, the shape of the "masking threshold pyramid" in Example 12 varies according to the masker level. [62] The relation of the masker level to the masked threshold is nonlinear; an increment in the masker level of 1 dB can produce an increment in the masked threshold up to 6 dB.[63]

---

[58] Fastl and Zwicker 2007, p. 64.

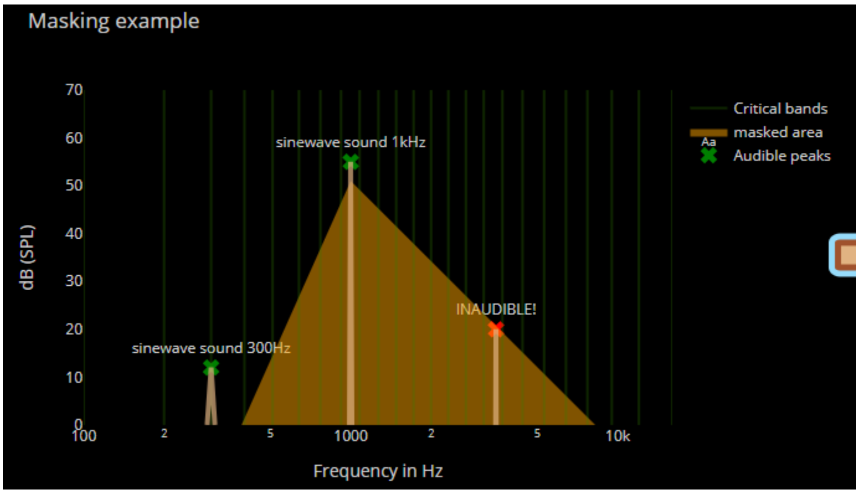
[59] Fastl and Zwicker 2007, p. 67.

[60] Fastl and Zwicker 2007, p. 67.

[61] Fastl and Zwicker 2007, p. 67.

[62] Fastl and Zwicker 2007, p. 69.

[63] Fastl and Zwicker 2007, p. 70.



Example 12. A simple masking example. When three concurrent sinewave sounds are playing, the highest is easily masked when the masking effect spreads from the lower sinewave excitation upwards.

5.8 MASKING OF COMPLEX TONES

In orchestral music, all tones have an overtone structure. Instrument sounds are composed of a fundamental tone, harmonics and low-level noise, and non-harmonic components. For complex tones, the masking threshold is determined by combining the masking thresholds of each sinusoidal component with their spreading frequency slopes. Therefore, the masking effect of an orchestral instrument depends on its spectrum. For example, a flute produces a strong single component and a faint spectrum, but a trumpet produces many harmonics, thereby creating a wide spectrum. Thus, a trumpet creates a broader masking effect than a flute. Tests show that when there are at least five sinusoidal components at the same critical band, the masking effect is like narrow band noise. At levels below 50 dB, even three components are enough to produce the same effect.[64]

When determining the masking curve of complex tones, it is important to remember that difference tones also produce masking. It can, however, be assumed that the frequency selectivity of the ear remains the same, irrespective of whether a complex tone or narrow-band noise is used as the masker. The difference is that complex tones, due to differential tones, produce more masking on the low-frequency side than narrow band noise.[65]

The sound of an acoustical instrument also varies in time. The high partials fade relatively quickly in comparison with the fundamental and low partials. The onset of an instrument sound contains a short burst of noise and non-harmonic components, which decay in milliseconds. It is then logical to assume that the masking effect of an orchestral instrument is strong at onset and fades along with the higher partials. In orchestral music, the amount of reverberation also affects the masking threshold, as onset components sound longer in a reverberant space than in a dry space.

---

[64] Fastl and Zwicker 2007, p. 72.

[65] Fastl and Zwicker 2007, p. 74.

When multiple orchestral instruments play simultaneously and the masking level of a single instrument is to be estimated, the question comes up of how to sum up overlapping masking thresholds. It is important to remember that the addition of masking does not follow the rules of the addition of intensity, where the addition is done inside the logarithm. The presence of two maskers does not automatically result in a 6 dB gain in the masking threshold, which would be the case of summing up two intensities of sine waves in the same phase.[66] Tests show that the strongest masker seems to be dominant, and additional maskers add only a little to the masking threshold. In one example, when four maskers produce the same masked threshold, 20 dB above the threshold in quiet, i.e., the level of hearing, an increment of only up to 21 dB can be achieved when the four maskers are presented simultaneously.[67]

In part of his virtual pitch theory, Ernst Terhardt used a simple formula to determine the mutual masking of spectral components of complex sounds.[68] Terhardt used this formula to reduce the number of a sound’s significant overtones to just a few important ones.[69] As a result, using Terhardt’s formula, there are usually 2-4 spectral peaks that define the formation of the virtual pitch.

In orchestral music, there is currently very little research on the masking phenomenon. In Part I, Chapter 8, the masking effect of orchestral instrument sounds will be discussed. According to one study, from a technical point of view, when two instruments are playing in the same range, one will not cover the other unless the sound pressure level difference is between 20-25 dB.[70] This conclusion, however, is based on notated tones and does not take the spectral content into account. The author of the study also points out that the audibility of instruments’ formant areas is essential for distinguishing an instrument in the orchestration.[71]

5.9 TERHARDT’S VIRTUAL PITCH

The complex harmonic or inharmonic sounds that consist of overtones are not perceived as chords, but as a single pitch. Some people can identify individual spectral components by listening, but in nearly every harmonic sound there is a pitch which is the base or root of the sounding entity. This root pitch is not necessarily present in the spectrum of the sound, but is something our brain adds because it fits into the existing spectral components.[72] Because the existence of such a root pitch cannot be measured in some cases, it is called the virtual pitch.

The Score-Tool App calculates the virtual pitch candidates according to Terhardt’s formula and shows them on a musical staff in an analytical view of the orchestration. From this data

---

[66] The information on summing up dB values and the effect of phase can be found in basic acoustical literature, such as Pulkki and Karjalainen 2015.

[67] Fastl and Zwicker 2007, p. 104.

[68] See Part I, section 5.9, “Terhardt’s Virtual Pitch.”

[69] Terhardt, Stoll, and Seewann 1982.

[70] Reuter 1996.

[71] The formant area is a concept borrowed from speech analysis and means the areas in the sound spectrum where the intensity of the partials is enhanced, regardless of the fundamental pitch. In order to determine audibility, the masking threshold at the formant areas weighs more than the masking threshold in general.

[72] Terhardt 1979.

one can see that, for nearly every instrument, the first candidate for the virtual pitch is the notated pitch. This means that we perceive music approximately in the same frequency region as if it were played by pure sine waves. However, because the sounding pitch is virtual, masking the sounding spectral pitches makes the virtual pitch disappear. Therefore, the masking phenomenon of the individual spectral components is more important than the masking of the virtual pitch. Of course, the virtual pitch can be, and often is, one of the spectral pitches.

To calculate the virtual pitch candidates, Terhardt introduced an algorithm that takes the mutual masking of individual spectral components into account. Combining that information with the idea of the spectral dominant region, Terhardt's algorithm points out the exact spectral peaks of the sound that are the most prominent in creating the virtual pitch illusion.[73]

In orchestration, virtual pitch appears in at least two different hierarchies. There is the collection of virtual pitches created by all the instruments of the orchestration chord, and then there is the virtual pitch which is the root pitch for the whole orchestration complex. Because the virtual pitches of individual instruments correspond to the notated pitch, I implemented in the Score-Tool App only the calculation of the virtual pitch for the whole orchestration.

5.10 SPECTRAL DOMINANT REGION

The sensitivity of our hearing system is described in the Loudness section (5.3 above). In equal-loudness contours, one can see that the most sensitive frequency area is at the 3.5 kHz point. However, when it comes to speech and music, this is not the area where the most relevant musical or speech information occurs. In a speech intelligibility test, it was shown that the biggest decrease in intelligibility comes when the frequency region around the 700 Hz area is cut from the audio.[74] It is no coincidence that the area is the same as where the first formants of vowels are measured.[75]

While I stated earlier that orchestral music often peaks at 300 Hz, 700 Hz appears to be another important frequency area in music, one that Terhardt calls the *spectral dominant region*.[76] One proof of its importance might be that many romantic composers often notated melodies for the piano in that area. On a piano keyboard, the 700 Hz area corresponds to the fifth octave, i.e., two octaves above middle C. In orchestral and vocal music, the same area is the optimal register for many instruments, from soprano singers to violin, oboe, clarinet, trumpet, and others.

I use the concept of a spectral dominant region in the Score-Tool App to determine the most important spectral peaks in instrumental sounds. For weighting the importance of peaks, I use Terhardt's formula, which is part of his virtual pitch algorithm.[77]

---

[73] Terhardt, Stoll, and Seewann 1982.

[74] French and Steinberg 1947.

[75] Pulkki and Karjalainen 2015, p. 87.

[76] Terhardt, Stoll, and Seewann 1982.

[77] Terhardt, Stoll, and Seewann 1982.

5.11 INFORMATIONAL MASKING

A special case of masking is *informational masking*. According to the definition of the term, it is masking that occurs beyond that which can be attributed to energetic masking.[78] A large part of energetic masking occur when the two tones excite the same region on the basilar membrane. In other words, a sound could be masked even if there were no other stimuli within the same critical band. Informational masking occurs when a harmonic sound with a simple spectral content appears with a sound having a rich spectral content, even with non-overlapping spectra. When listening tests were made using a single sinusoid against a complex harmonic sound, informational masking was strongest, masking even 50 dB sinusoids, when the spectral content of the masker was fewer than 20 partials,[79] which applies to most of the orchestral instruments.

Interestingly, when the sound of the informational masker has a high number of partials, meaning over 20, our hearing system treats it as noise. As a result, the normal rules of auditory masking apply. In orchestral music, informational masking is thus at its strongest when a sinusoid-like instrument, for example, a recorder, plays against a small ensemble, for example, a wind quartet. The informational masking concept is currently not implemented in the Score-Tool App.

6 ORCHESTRATION IN THE DIGITAL ERA: CAN TECHNIQUES USED BY RECORDING ENGINEERS BE APPLIED TO ACOUSTIC ORCHESTRATION?

In this chapter, I discuss similarities between the tasks of the recording engineer and the orchestrator, mainly from the perspective of what I, as an orchestral composer, can learn from a recording engineer's work. The recording engineer has become an essential part of the music industry, and while their workflow was little discussed in literature before the millennium, the growing number of home studios has raised the interest of music consumers in the recording engineer's knowledge.

6.1 THE RECORDING ENGINEER AS ORCHESTRATOR

The mixing of orchestral music, similar to the work of composing, has been somewhat mystified as a task that requires special skills. Mixing differs from orchestration in that it is mainly done by ear, relying on subjective perception. The interesting thing is that, to make a good mix, engineers intentionally use not-so-good equipment in order to hear how the mix will sound outside a studio environment.[80]

Good recording engineers are in demand in today's music industry, because they are the ones who finish the orchestration for musicians, an issue I discuss later in this chapter. In the context of popular music, a poor mix can ruin even the best music, and a good mix can make an average piece sound good. From an orchestral composer's point of view, a poorly orchestrated passage certainly does not bring out the best in a composition, while a poor composition orchestrated masterfully can sound amazing.

---

[78] Kidd, Mason, Richards, et al. 2008, p. 144.

[79] Gelfand 2009.

[80] This might also be a useful way for composers to experience how the orchestration would sound played by amateurs.

Recording engineers can be seen as modern-day orchestrators, at least in the field of popular music. The difference from acoustic orchestration is that these engineers get immediate feedback on their work by listening to the mix. A composer, on the other hand, must rely on known good practices, orchestration handbooks, word of mouth, and their instincts while orchestrating, and feedback often comes all too late – in orchestral rehearsals. With the Score-Tool App, it is possible to get feedback on one's orchestration in the same way recording engineers can listen to their mix.

Mixing also permits a trial-and-error type of workflow because the result is not expected to be published before errors are corrected. The composer's work, on the other hand, is based on trial and not admitting errors, because it is not wise to admit that an artist is not a master of one's own field. The Score-Tool App gives a composer the opportunity to have a trial-and-error type of workflow, because the errors do not become public. It is thus useful to look at how the trial-and-error-based workflow could fit into the work of composition.

6.2 MY ARTISTIC INTEREST IN THE RECORDING ENGINEER'S WORK

In this chapter, I refer to my discussion partners anonymously as recording and mix engineers.

In the year 1999, I had my first orchestral piece recorded on a commercial CD. I remember that in the recorded work, *Tears of Ludovico,* there were numerous audibility issues which I noticed in the rehearsals and the performance. One unsolved problem was the weak sound from the harp in the middle and low registers, where the instrument was masked by almost any simultaneously sounding instrument in the orchestra. I was first irritated by my obvious orchestration mistake, but then I received help from an unexpected side: the recording engineer. He heard my dissatisfaction and agreed to bring up the harp volume in the final mix of the CD release.

At that time, I was understandably relieved by the possibility of adjusting the orchestration without trial-and-error experiments, which would have involved altering the orchestral parts and having cumbersome communications in a foreign language with the conductor and the orchestral players. Further discussions with the recording engineer revealed that this kind of touch-up is common in orchestral recordings. Especially in solo concertos and in orchestral music with voices, the audibility of the solo part is enhanced at the mixing desk, and the orchestral balance is not equivalent to the acoustical one. Even in compositions without a soloist, the balance is adjusted afterwards, because the sounds of individual instruments are captured with microphones placed near the sound source and not only with microphones in the audience. There are, however, recordings that are intentionally made with just a stereo pair of microphones placed among the best audience seats.

Based on my personal discussions with recording engineers, in orchestral recordings with individual microphones on instruments the final mix is done by respecting the acoustic character of the original instrument. This means that the spectrum of the original sound remains as close as possible to its acoustic equivalent, and the mix is done mainly by adjusting the levels of a sound's power. It is usual to cut the frequency areas outside an instrument's range, especially the lower range, so that unwanted noises and microphone characteristics in those areas would not disturb the overall balance. For example, the lowest fundamental frequency of the flute is in the 260 Hz area, so it would be safe to filter out the frequencies below 200 Hz.

In the case of the harp, which was my original concern in the *Tears of Ludovico,* the frequency bands were not cut, because the instrument's range is wide. Raising the level of the harp's sound in the mix results in raising the levels of the player's unwanted stomps and sighs as well, which could have been removed by adjusting the frequency curve, i.e., applying equalization or EQ. The recording engineer was not willing to apply EQ to orchestral music because it would have altered the original sound.

After this initial contact with a recording engineer working with orchestral music, I have tried to have discussions about the orchestration with the studio people whenever possible. Since 1999, I have had several orchestral works recorded on CD, all of which required a great many tweaks and adjustments in the balance. The engineers sometimes seem to be protecting their professional expertise and do not reveal what they are actually doing when they raise the sound of an instrument, yet the end result has nearly always been satisfactory. However, the same people are always willing to discuss my orchestration choices and the effect that I am seeking with a certain kind of instrumental combination in order to achieve that effect in the mix.

I have also had interesting discussions with a mixing engineer who did the sound for one of my operas, which was performed outdoors and thus was amplified. Every single instrument in the orchestra and every single singer, even those in the choir, had a microphone, and the orchestration balance was adjusted on the mixing board. This was also an opportunity to think about my general preference in orchestral sound. The end result was a mix that sounded as if the music was performed in a good hall with only a slight reverb and as if the singer were right in front of us with the orchestra several meters away. This made me think about my orchestration ideals; if this is my preference, do I subconsciously try to orchestrate in a way that the music would sound like that? At least in operas, this might be the ideal balance, since even Wagner designed the orchestra pit at Bayreuth’s Festspielhaus to have a hood which corrects the volume balance between the orchestra and the singers; the orchestral sound is attenuated while the singers' voices are not.[81] The good thing in the outdoor opera mix was that I did not feel the need to shape the spectrum of any sound, and neither did the engineer.

My most important encounter with a mixing engineer was in 2014, an encounter that also initiated the Score-Tool project. That year I composed a 45-minute long work titled *1900 – Virtual Piano Concerto*, for solo piano and loudspeaker orchestra. In that piece, I used pre-recorded orchestral parts played back from an array of loudspeakers mimicking the directionality of sounds as these come from the seating arrangement of a symphony orchestra.[82]

In that work, I had practically unlimited possibilities for tampering with the orchestration balance because every instrument was played back from its own channel and its own speaker. Still, no matter how I adjusted the balance and physically turned the speakers to mimic the directionality of the real instruments, the orchestra did not sound authentic to my ears. In one performance, there was a mixing engineer in the audience, who came to thank me after the concert. The first thing he said was, “You should have used the EQ.” “How?” I asked. “To clean up the tracks,” he said.

For that project, the EQ task was too great to carry out between concert days. Afterwards, I tried to insert a few EQ sections in the piece and noticed that the orchestra actually sounded more authentic after the adjustments. This raises an interesting question: why do recorded orchestral instruments sound more authentic after cleaning up the unwanted frequencies in a recording? Part of the answer is that the effects caused by the recording equipment are

---

[81] Barron 2009, p. 352.

[82] The performances were realized in collaboration with Tapio Lokki and Jukka Pätynen of the acoustics department of Aalto University (in greater Helsinki).

eliminated, and also in sample recordings, many auxiliary noises that are also captured and amplified by close micing aren't usually heard as distinctly at adistance from the instrumentalist. But in my view, there is more to it than that. Similar to the reverberation issue, which will be discussed in Chapter 10, perhaps our hearing system prefers the cleaned-up version of an orchestral recording, and the untouched recording would be our second-best choice.

6.3 WHAT AN ORCHESTRAL COMPOSER CAN LEARN FROM A RECORDING ENGINEER

There are big differences between the work of recording engineers in orchestral music and those working with popular music. For example, in my experience with orchestral music, the balance between instruments is thought to be more sacred in comparison with popular music. By this I mean that in orchestral recordings, soft instruments should sound soft, but in popular music, for example, the low strings on the acoustic guitar can be intentionally mixed as loud sounds. Here, I discuss the work of popular music recording engineers on the basis of available literature. I also seek the recording engineer's equivalent actions in orchestration practice and look for ways in which my Score-Tool App can help in this workflow.

In recording electronic and amplified music, the approach to mixing is different than is the case with acoustic music, because the recording engineer does not have to be concerned about staying true to the original sound.[83] There is no such thing. The balance of the mix can therefore be adjusted by shaping the frequency curve of the timbre. In mainstream popular music, it is usual to place each instrument in its own frequency area utilizing the critical band structure of our hearing system. In this way, the power levels of each instrument can be adjusted freely because they are not competing with each other for dominance in loudness. Although this technique is seldom used in recordings of orchestral music, there are principles in the engineering practice in popular music that may stimulate new ideas even in the field of acoustic orchestration. In the end, the underlying problem is the same: what can we do to make inaudible sounds audible?

In the field of analytical acoustics, the masking phenomenon with complex acoustic tones has been tested very little because of the difficulty of controlling the large number of variables. The same phenomenon, however, is tested indirectly by mixing engineers whenever they create a mix from a multitrack musical recording. The task is often to reduce the mutual mask of the tracks, which happens when multiple tracks share the same critical band. When music is imbued with a large amount of information, the segregation of the instruments has been shown to improve only when spectral components of the loudest instrument are left on each critical band.[84] A test using this method was conducted with 11 music technology students mixing seven cases, but the result was not uniform. Nevertheless, the method was preferred, not all, but by approximately two-thirds of the test participants. This is good news for orchestral composers, because cutting frequencies entirely from a specific band is not possible in acoustic music. Because of the ambiguous results of the cutting method, there are other ways to make

---

[83] Juth 2021, p. 10.

[84] Kleczkowski, Plewa, and Pluta 2011.

desired instruments audible in the mix. Three common methods for reducing masking are mirrored equalization, frequency spectrum sharing, and stereo panning.[85]

In mirrored equalization, the spectral content, i.e., the frequency region, is attenuated for one track in the whole mix and boosted for another track which shares the critical band. This helps the boosted track to be heard with very little alteration at the level of each track. The orchestration equivalent would be to use a *sordino* or change the playing technique so that it affects a certain frequency area. The important step from the mix engineer's viewpoint is not only to boost, but also to cut in order to maintain the overall loudness.

In frequency spectrum sharing, the method resembles mirrored equalization, but the actions are more radical. One track involves high-pass filtering, while the other involves low-pass filtering. "Pass filter" means that only certain frequencies are passed, so this method places each pair of tracks at the far end of the spectrum, ensuring that their spectra do not overlap. The milder version of the technique uses only a high-pass filter or only a low-pass filter on one track. The orchestration equivalent would be registral positioning of the musical material so that there is minimal overlap of each material's spectral content. The Score-Tool App is ideal for minimizing the overlap because the frequency content of the orchestration is shown graphically.

Stereo panning means just what its name says – panning the tracks that share the critical band to different places along a left-right axis in stereo image. This can be done either by panning both tracks to the far ends or by keeping one in the center and panning the other.

These three methods can give added clarity to the sound, but recording engineers are sometimes expected to give something extra, which must be frustrating to them since they cannot boost anything that is not already present in the music. The same thing applies to the orchestration. The problems in the overall timbre of a piece may not always be an orchestration problem; the problem may lie in the composition itself. This is also the view of some recording engineers, namely, that the original material should be recorded in such a way that there is very little need for tampering with the frequency boosts and cuts. This can be done, for example, by moving the position of the microphone or the position of the performer.[86] In an orchestral composition, the composition should be written with such choices that a good orchestration balance can be achieved with minor tweaks to the dynamics or register, and not with radical operations, such as leaving out instruments or by asking everybody but the soloist to play *pianissimo*.

The recording engineers' view is that there is a direct relationship between the need for EQ and the number of instruments in the mix. If the music consists only of a basic rock combo – guitar, bass, drums, and vocals – then very little EQ is needed in mixing the piece. In a large arrangement with backing vocals, synthesizers, and horns, the mix needs to be managed by a skillful recording engineer.[87]

In my view, the development of mixing practice is one of the reasons for poor orchestration in recorded music, because prior to having the ability to manipulate frequency, the only solution for masking problems was to orchestrate well, even for popular music. With a recording engineer present, the balance of popular music's orchestration is placed in the engineer's hands

---

[85] Wakefield and Dewey 2015.

[86] Benediktsson 2019, p. 58.

[87] Benediktsson 2019, p. 53.

whose task is to solve “bad” orchestration with electronic tricks. This was exactly the case with my inaudible harp, and I accepted the help light-heartedly then. Afterwards, I was frightened by the consequences of the solution, having mainly relied on the engineer as the savior for bad orchestration.

In popular music there are some common well-known cases which results the masking problems in the mix, such as bass guitar and kick drum, rhythm and lead electric guitar, electric guitar and synthesizer etc.[88] These instrument combinations nearly always require actions other than altering the overall power level of the track. In popular music it is also common for string sounds to drown the piano, and literally everything interferes with the vocal part. These kind of cases are corrected on the mixboard.[89]

The last argument will be familiar to composers of vocal music, because audibility issues are commonplace in music in which the full orchestra plays along with a vocal soloist. The orchestration point of view on this matter is discussed in Part I, Chapter 3 above, but here I add that the vocal part’s sensitivity to masking is higher than an instrument’s. The reason might be that there is a subconscious need to hear fully the person who is singing. Very few of us like to have a discussion with a friend in a loud environment. That’s why orchestral passages with vocals must be orchestrated with extra care, so that the orchestral part does not disturb the vocal part, but instead supports it. This also applies to any large ensemble with any soloist, but the vocal part is perhaps the most sensitive to masking issues. Questionable advice from recording engineers is that the more instruments you have, the better EQ skills you need.[90] From my perspective, the more instruments you have, the better orchestration skills you need.

There are a few interesting cases where frequency masking is a desirable effect and is created intentionally. In popular music recordings where the vocal technique of singers differs from classically trained singers, there is often an unwanted nasal vocal sound, which comes from a combination of a certain kind of voice-and-recording technique. According to mix engineers, the effect lives at the 1-1.2 kHz area in the frequency spectrum.[91] Attenuating this area helps reduce the sound’s nasal quality, but sometimes the attenuation affects the vocal sound body, making it sound thin. In these cases, the nasal area could be intentionally masked so that the nasality becomes inaudible while the body of the sound remains untouched. The same technique can be used in the orchestration, where the nasal quality is attributed to reed instruments, such as the oboe, English horn, and bassoon. In theory, the nasality could be masked by adding instruments with a strong fundament at the C5-C6 octave area, corresponding to the mix which engineers have defined as the frequency area for nasal sound. The choice of instrument to add can be made with the Score-Tool App.

6.4 INSTRUMENT PLACEMENT IN THE HALL AND AUDIBILITY

We have seen that two methods used by recording engineers – mirrored equalization and frequency-spectrum sharing techniques – can be translated into orchestration practice and tested and pre-evaluated with the Score-Tool App. In this section, I discuss a third method,

---

[88] Wakefield and Dewey 2015.

[89] Benediktsson 2019, p. 54.

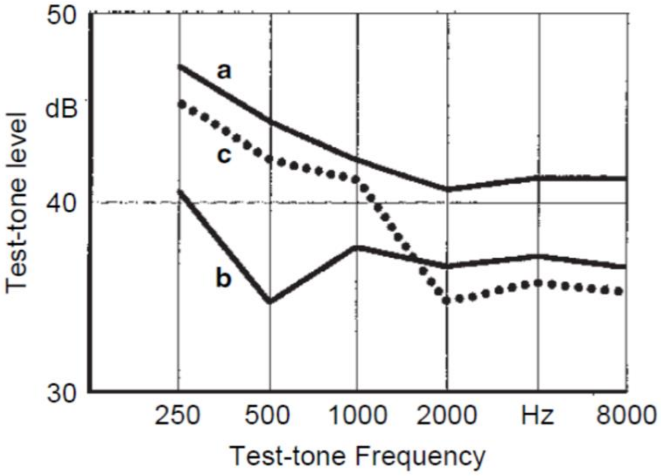
[90] Personal conversation with an experienced mix engineer in 2021.

[91] Benediktsson 2019, p. 52.

stereo panning, the difficulties of translating this method into orchestration practice, and a test for the best way to make a given instrument audible in the orchestral mix.

EQ is not the only method used to solve audibility problems, at least for recorded music, where, in theory, sound can be sent to only one ear of a human's two. Stereo panning affects frequency masking when the masker and the maskee sounds come to separate ears. The panning technique relies on the individuality of our two inner ears: the masking phenomenon in one ear does not entirely mask the frequencies in the other ear. Despite the fact that we actually have two distinct hearing systems, the data coming from both systems mix together in our brains. According to various studies, the ears are not independent. A sound in one ear affects the frequency detectability in the other ear. This is called binaural interference, which also involves masking. But despite the interference, in stereo the masking effect is somewhat reduced.[92]

Our perception of sound is highly sensitive to the direction from which it comes. Even a slight change in the position of a sound source, including from behind us, is somewhat detectable.[93] The directionality of the sound source also affects masking. When a test sound and a masking sound reach a listener from different directions, the masking is not as strong as when the sounds come from the same direction.[94] This phenomenon is, as always, more complex than simply measuring the position of the sound source. In Example 13, borrowed from Jürgen Meyer's *Acoustics and the Performance of Music*, a graph shows how the directionality of the sound affects masking in the function of the frequency of the sound.[95]



Example 13. The graph, in Meyer 2009, p. 17, shows the effect of reduced masking when the masker and maskee come from different directions. The masker is pink noise and the maskee is a sinusoidal test-tone. *Curve a* is a sinusoidal tone and noise that come toward the listener directly from in front. *Curve b* is a tone that comes toward the listener 60 degrees from the side

---

[92] Moore 2012, p. 259.

[93] Pulkki and Karjalainen 2015, p. 238.

[94] Meyer 2009, p. 16.

[95] Meyer 2009, p. 17.

> at the same time as noise comes from the front. *Curve c* is a tone that comes toward the listener 60 degrees from the side with noise as a diffuse sound field.[96]

The graph in Example 13 shows how noise masks a sinusoidal tone when the position of both is altered. The noise position is changed from directly in front (a and b) to diffuse (c). The tone position is changed from directly in front (a) to the side (b and c). The concert hall circumstances match the diffuse sound field more closely than the directional sound, which is discussed in greater detail in Part I, Chapter 10 ("Another Actor-funded project related to my interests is the OrchView software, which at the time of this writing is under development. OrchView allows users to annotate scores according to their perception of music. The annotations are performed by analysts listening to specific performances and looking at scores. The annotated scores are gathered into a database, which is called ORCHard, another project in the Actor community. The ORCHard project currently has a very large database of excerpts from scores of well-known orchestral works with tagged features, such as blending chords, contrasting chords, background texture, and so on. All excerpts also have listening examples from historical recordings, produced in various halls without touch-up of the dynamics. The database is searchable and is maintained by an orchestration researcher, Stephen McAdams. The database is semi-open; it is accessible to anyone with a user account, and the user account can be requested from the database webpage. The database could be used to determine the audibility of an instrument by searching the equivalent orchestration with appropriate tags and comparing that to your own. This technique is similar to that used in orchestration handbooks. From my perspective, the database can be used as test case, for example, to test passages tagged as blending chords and to try and determine with the Score-Tool App if blending is suggested in these same passages.

The effect of the concert hall on the audibility of an instrument"). The graphs show that masking is most significantly reduced overall and especially at the 500 Hz area when both sound sources are directional and the target is from the side, and masking is reduced only at about 2 kHz and up when the masker is diffuse.

In orchestral music, from the point of view of my Score-Tool project, both the masker and the target are complex sounds. In addition, the web of parameters affects the masking and becomes so complex that testing becomes difficult. One more thing that factors into this question is that as listeners, we do not keep our heads still while listening, but we look around at the orchestra, focus on the players, and turn our heads in different ways. However, when multiple sound sources are distributed around a listener, the hearing mechanism (inclusive of further information processing in the brain) has the ability to concentrate selectively on one of these sources and emphasize it in comparison to the others. This phenomenon is referred to as the "cocktail party effect."[97]

The cocktail party effect is a combination of directional hearing and psychological aspects and is therefore beyond the scope of the Score-Tool project. The important thing in defense of my decision to concentrate on monaural masking is that tests on the cocktail party effect show that the sound pressure level of the sound of interest lies about 10–15 dB above the masking level determined by the masking sound. Otherwise, directional location is no longer possible. [98] In

---

[96] Example from Meyer 2009 after Prante et al., 1990.

[97] Meyer 2009, p. 17.

[98] Blauert 1974, p. 4.

other words, the cocktail party effect does not cancel frequency masking, but it does reduce informational masking.[99]

The cocktail party effect in music has been tested under a different name and in tests more thorough than the one in Meyer's book, conducted with a loudspeaker array around the test subject and playing noise from one side and a sine wave signal to be detected from other side. In an anechoic chamber, a masking threshold up to 18 dB lower was detected in comparison to the same sounds played from a single source. This phenomenon is called *unmasking*.[100] The orchestration equivalent to unmasking would require placing the conflicting instruments on opposite sides of the hall, although as I stated earlier, this is more of an artistic statement than a masking solution.

Unmasking in general depends heavily on the listening conditions and applies fully only to listening through headphones. With loudspeakers, there is a cross-talk phenomenon, whereby the sound from one speaker reaches both ears, although the sound is localized to the speaker in question.[101] Cross-talk also happens, of course, in acoustic concerts. Furthermore, sonic reflections in the hall disturb the live sound localization.

With regard to panning in orchestration, the audience's distance from the instruments moderates the effect, because the stage might be only 15-20 degrees wide in our hearing width of the full 180 degrees possible in headphone listening. With headphones and the binaural psychoacoustic methods, it is possible to go even wider and put the sound source in an arbitrary place in a 3D field. Of course, it is also possible to place the players behind the audience in a concert hall, but in my view that is a special artistic statement rather than a solution to audibility problems.

The effect that panning has on frequency masking is not to be overlooked. Proof of its effectiveness was shown in a test in which the test subjects had to adjust the intended target track to be audible from the mix. The adjusting was done by rotating an unlabeled knob; i.e., the test subjects did not know the knob's function. The test showed some interesting results. Of the three masking reduce methods discussed earlier, some recording engineers favor the mirrored equalization technique, but this was the least favored in the test. The test subjects preferred the stereo panning method.[102]

The result, translated to orchestration practice, means that the best solution to audibility problems might be placing the inaudible player to the extreme left or extreme right of the stage. This, of course, makes questionable the convention of placing an orchestra soloist in the middle of the stage. The effect that stage position has on panning, as stated before, is quite subtle compared to the panning knob on the mixing board. However, in my work *Sonority*, discussed later in Part II, Chapter 3 ("*Sonority* – horn concerto, November 2, 2021"), the soloist is placed on the far left and then on the far right side of the stage at the beginning and the end of the piece respectively, and I noticed increased soloist audibility in those passages from where I was listening in the audience.

---

[99] For further information, see Part I, section 5.5.

[100] Pulkki and Karjalainen 2015, p. 243.

[101] Pulkki and Karjalainen 2015, p. 294.

[102] Wakefield and Dewey 2015.

As stated earlier, currently the Score-Tool App does not take spatial separation into account in masking calculations. The reason is the uncertainty in the panning sensation that actually reaches the audience in the hall and the trouble of putting this uncertainty in an algorithm. However, the effect of spatial separation and directivity of orchestral instruments on masking might be a good subject for further study, as my experience with my composition *Sonority* also shows.

6.5 CONCLUSION OF THE DIGITAL ORCHESTRATION CHAPTER

In conclusion, the optimal solution for making an instrument audible in orchestration with an engineer's mixing method might be to apply all three techniques mentioned in this chapter where possible:

- Use the different playing techniques or *sordino* to alter the sound spectrum, and check the effect on the orchestration with the Score-Tool App.
- Use registral separation as the target. The effect of registral separation can be affirmed with the Score-Tool App.
- Use the spatial separation on stage for competing instruments. The effect of spatial separation is currently not implemented in the Score-Tool App.

These three methods, effectively used, should be enough to solve orchestration audibility problems if they are solvable. In the original case with the harp, the solutions might be ambiguous.

7 AUDIBILITY AND BLENDING OF MUSICAL INSTRUMENTS

7.1 PREVIOUS RESEARCH ON BLENDING MUSICAL SOUNDS

The blending of different instrumental timbres is a fuzzy subject, and the term "blend" is used both in negative and positive instances. Blending timbres is somewhat related to mixing timbres, but in my view, mixing is something that happens automatically, whereas blending is a special case of mixing in which timbres mix particularly well. A comparison would be a mix of oil and coffee. We could say that they mix, but do not blend; the mix will be a dark liquid with splotches of oil. However, if we mix milk and coffee, we get a uniform blend of light-brown liquid.

Blending timbres is not only related to the timbre, but also to the harmony and attack time.[103] Also, blend affects timbre because time depends on concurrent grouping.[104] In this report, I leave attack time out of the calculations because my focus in this thesis is not on temporal changes of the sound.

Gregory Sandell states that blended musical combinations are those in which the distinctiveness or individuality of the constituent instruments is subordinated to obtaining an overall, uniform timbral quality.[105] In his definition, the distinctive timbre of both instruments will be lost, and the resulting composite will sound as though it originated from a single source. This definition is in my interest, since in this kind of blend, the instrumental sound "gets lost" without masking. That being said, it is much more unlikely for sound to blend unintentionally

---

[103] Tardieu and McAdams 2012.

[104] McAdams 2019, p. 211.

[105] Sandell 1991, p. 40.

than to be masked unintentionally. Blending is included in my research as a special case of the inaudibility of the target instrument, which a composer can avoid in an unwanted case or enhance when desired.

7.2 BLENDING AND HOMOGENEITY OF THE TIMBRE

Blending is to some extent linked to the homogeneity of the orchestration. In combining instrumental timbres in chords, the result can be said to be heterogeneous if the chord is composed of sounds with contrasting timbres, and homogeneous with matching timbres. Contrasting timbres are less likely to form a blended chord than matching timbres. An example of this would be a 5-note chord from each orchestral woodwind versus the same chord from each orchestral string instrument, of which the latter would be naturally homogeneous. It has also been said that the general rule for achieving blend is to consider the "affinity" of the instruments to be combined. Although not explicitly defined, affinity among instruments implies a perceived similarity.[106]

There is not, however, unanimous opinion on whether the blend requires two or more separate sound sources to be distinctive or matching in their spectral content. The technique of creating so-called *super instruments* from even contrasting timbres includes the idea that the resulting sound would be perceived as if it were a kind of rich organ registration, i.e., as a single instrument rather than a combination of many. Examples of super instruments include a well-known passage from Ravel's *Bolero*, where towards the end, the melody is orchestrated with doublings in parallel dominant seventh chords. Doublings include instruments from different groups, even those contrasting in timbre, but the resulting sound is perceived as a blended super instrument.

There is no general guidance for creating these kinds of super instruments, and there are no explicit studies that predict which instrumental timbres are likely to be dominant in complex orchestrations.[107]

7.3 BLENDING AND INSTRUMENT FORMANT AREAS

The opposite view comes from Christoph Reuter, who divides simultaneous timbres into *blended* and *partial masked* categories. The blended category requires the timbres to have equivalent formant areas, i.e., the spectral content, and requires the partial masked category timbres to have non-matching formant areas.[108] His hypothesis is that, in the partial masked category, all timbre components are perceived separately. The result was in fact that the sounds with matching formant areas were perceived as a continuous "blended" sound while sounds with non-matching formant areas were not. Similarly, traditional orchestration handbooks speak of the combination of timbre-matching instruments as a "good combination."[109] In my view, the combinations of matching timbres can result in an unwanted blend, but for non-matching timbres to blend, the composer or orchestrator must create this intentionally; blending does not happen by accident.

---

[106] Goodchild and McAdams 2018, p. 500.

[107] Dolan and Rehding 2021, p. 502.

[108] Reuter 2000, p. 1.

[109] Rimsky-Korsakov 2013, p. 117.

7.4 THE HORN AS A GOOD BLENDER

Throughout orchestration treatises, the horn sound is regarded as a timbre that blends well with all other instrumental timbres. Pointing out a single instrument that has a blending property indicates that the timbre similarity is not necessarily needed, but different timbres blend if at least one of them is “blendable.” This was the case with my work *Sonority*, where the timbre of the solo horn tended to blend in the orchestra even in passages where the soloist’s sound was definitely not masked by the orchestration.

The blending ability of the horn can be found in numerous orchestral scores, where the horn often acts as a pedal, echoing harmonic key tones in the background, obviously without the intention of the horn timbre being distinguishable. The pedal tones are, specifically in the classical era, orchestrated in octave doublings with two horns.

An interesting point of view found in orchestration treatises is to consider the horn timbre as the glue that results in the blending of different instrumental groups. All four authors presented in the orchestration treatise chapter advise using the horn to link string and woodwind timbres, for example. There are also other instruments said to have this ability: the bassoon in the high register, string harmonics, and the saxophone – all instruments with vague mutual ground in terms of spectral content. The idea of spectral properties of the instrument’s acoustics as blending glue also indicates that the effect of blending depends on the relations among the spectra of the resulting mix, not only on the relations among the spectra of the individual components.

7.5 THE SPECTRAL CENTROID AND BLENDING

Several authors, including Sandell and Lembke, suggest that timbres with spectral content concentrated around the fundamental frequency have the ability to blend well. This applies especially to the horn timbre in the middle register, which has a strong first partial in its spectrum accompanied by only a few low amplitude overtones.

Another good test candidate for this hypothesis is the violin, with or without the mute. Since the mute affects the violin timbre by damping the higher partials, muting should, according to this hypothesis, make the sound more blendable in comparison with the unmuted violin. There is in fact a mention of this kind of behavior by jazz orchestrator Nelson Riddle, cited by Sandell, who in his study of arranging observes that one of the wonderful aspects of mutes, the blend, whether it be a small section or a large one, improves magically with use.[110]

The number of high partials in the timbre can easily be expressed with one value, the *spectral centroid*, which describes the “center of mass” of the timbre. The spectral centroid is defined in its own section (Part II, section 4.7). If, for example, the timbre is thought of as a long wooden stick, the spectral centroid would be the point at which the stick is in perfect balance. A stick with a uniform distribution of mass would have the balance point in the center, and a stick with a metal handle would have the balance point near the handle. In similar fashion, the root tone of an instrumental timbre can be thought of as the handle and the last audible partial as the other end. Thus, if the timbre has a low spectral centroid, this automatically means that the timbre has weak upper partials, leading to good blending abilities. For example, Lembke

---

[110] Riddle 1985, p. 124

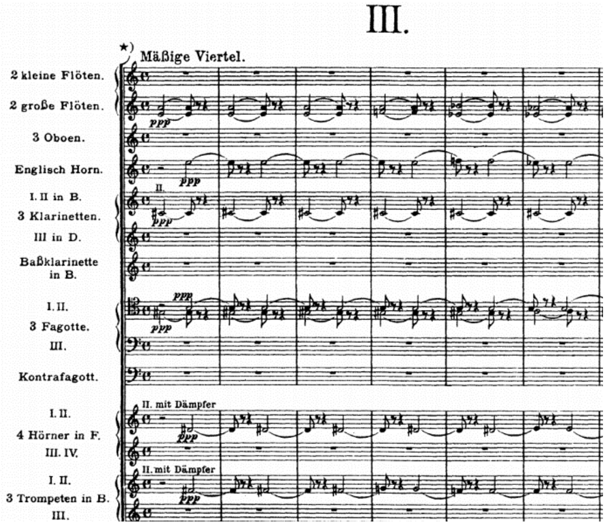
devotes a major part of his study to the blend of timbres with low centroids. He writes, “In conclusion, matching spectral features between instruments appears to be a general strategy to achieve blend […] In addition, there are several indications that a relative `darkening' of timbre is also understood as a general strategy […].”[111] The ‘darkening’ means low dynamics on the upper partials, thereby meaning low centroid.

Taking into account the effect of the spectral centroid along with my experiences with my *Sonority* concerto, I added a formula in the Score-Tool App to warn users about decreased audibility when the spectral centroid of the target is low.

7.6 THE BENEFIT OF LOW DYNAMICS TO BLENDING

So far, I have only discussed the spectral features of instrumental timbres in relation to the blend. Other important features, which are somewhat linked to the spectral features, are dynamic level and harmony. The dynamic level plays an intuitively large role in blending, as it is hard to imagine, for example, the sound of a *fortissimo* trombone blending into any other sound. In almost any acoustical instrument, the rise in the dynamic level also results in the rise of the spectral centroid. This can be seen, for example, using my Score-Tool App. The effect is due to the relative amplitude of the high partials, which is most evident in a brass instrumental group. The loud high partials give the timbre its signature “brassy” sound, which is described as having “penetrative quality” by all four authors discussed in the orchestration treatise chapter.

There is also a tendency among orchestration teachers to suggest low dynamics when a balanced-sounding orchestral chord is desired. In a *fortissimo tutti* chord, often at the end of the section or of the whole work, the blending capabilities of individual instrument sounds are seldom discussed. This tendency can be seen clearly in Schönberg’s orchestral work *Farben*, a piece centered around the idea of changing the color of orchestration. The beautiful blending of interlocking chords is achieved by keeping the dynamics low, even as low as *ppp*. The very nature of the piece suggests low dynamics, but it would be an interesting experiment to see if



Example 14. Arnold Schönberg, Five Pieces for Orchestra, Op. 16, no. 3, *Farben*, mm. 1-6. The blending of timbre is achieved with the help of low dynamic levels.

---

[111] Lembke 2015, p. 126.

the nature changes by raising the dynamics, and how much rise the texture would tolerate before losing its beauty.

Dynamic levels of blending orchestration can be thought of as analogous to blending colored lights or paints. It would not be the best idea to use blinding bright lights to try blending or to try blending paint by using neon colors. Blending, at least when done on purpose, calls for subtle changes in intensity and balance. This speaks once more against the idea that blending orchestral instrument timbres could happen by accident.

7.7 MUSICAL DISSONANCE AND BLENDING

While dynamics affect blending by heightening the spectral centroid in the spectral scale, harmony affects blending regardless of the spectral centroid. When a musical dissonance, such as a minor second, is playing even with two pure sinusoidal waves without any spectral content, the waves are likely to be perceived as two sounds instead of one blended entity. However, saying “two different sounds” would be to exaggerate, since with the interval of a minor second, both tones would lie inside the same auditory band, rendering the sensation of two tones impossible to perceive. However, the beating phenomenon created by two close wave frequencies reveals the interval, at least to a musician.

Generally, chords having tones with matching spectral components tend to blend better than chords having tones with non-matching components. This is evident in classical-romantic scores, where large chords are often placed in a register that imitates the natural overtone series. This makes many of the overtones of lower-register instruments automatically match overtones of the instruments above. It is, of course, understandable that the unisons and octaves work best in this manner, and those intervals are still widely used in twentieth-first-century music, when blended-sounding super instruments are created.

If complex harmony is used in orchestration, the colliding dissonant intervals, described above, disturb the blending sensation. The web of overtone partials of each tone becomes so dense that roughness and beating will inevitably be detectable on many auditory bands. However, if most of the partials are spread among several auditory bands, the listener can perhaps adapt to the situation, ignoring the roughness and beating and perceive the entity as a single timbre. This is exactly what happens in Schönberg’s *Farben*, where the composer uses a complex and dissonant harmony intended to blend as one entity. In *Farben*’s case, the repetition and static mood help to achieve the adaptation.

Harmony is also related to intonation. Since intonation that differs from natural overtone spectrum creates roughness and beating, it helps to discriminate among individual instruments which are not in tune within the chord and therefore diminish the blend. Intonation is a complex subject, and I will not go deeper into it here. In terms of audibility, however, intonation can be used to advantage. As a professional violinist once observed, in performing with an orchestra, intentionally playing slightly out of tune helps the violin to stand out from the orchestra.[112] This is, of course, a technique applied by a highly experienced musician with utmost care. I would not recommend that composers notate out-of-tune pitches in the score.

---

[112] Personal conversation, May 2013.

7.8 MUSICAL REGISTER AND BLENDING

In orchestration, not only do the notated pitches play a role in blending, but register also has a role. Tones close by in register, if not too dissonant, blend better than tones far apart. This is not always as straightforward as counting the octaves between doublings, because certain combinations, such as the high piccolo with a low trumpet can create a blended sound with sounding root tones even three octaves apart. The registral closeness becomes important in blending sounds like the trombone with the violin, which are hard to blend even in unison. The effect of musical register in blending can be interpreted from the graphs in the Score-Tool App. The detailed effect of musical register in blending could be one subject for future developments in the App.

7.9 CONCLUSION ABOUT PREVIOUS RESEARCH ON BLENDING

Both Sandell and Lembke come to the conclusion that the two most important timbral parameters supporting the blending sensation are darkness, i.e., the spectral centroid, and matching timbre. Reuter adds the importance of the performance parameters of micro-modulations and vibrato as well, but in my Score-Tool project, I decided not to count these and have concentrated instead on the timbre parameters introduced by Sandell and Lembke.

Comparing the instrument timbres using the spectral centroid is straightforward, because it is just a number value indicating the center of mass in Hertz. Comparing timbral features is, however, tricky. Timbral information can be read from the sound spectrum, but for that, the spectrum must be filtered somehow to include only the relevant information. For example, comparing the sound spectra, without a filter, of two different tones more than an octave apart on the clarinet does not immediately show that the tones are coming from the same instrument. Therefore, there are numerous algorithms intended to extract information for different purposes from the spectrum.

With one such algorithm, timbre or “sound color” can be represented in a quite compact form using *mel filter cepstrum coefficients* (MFCC). The power of MFCC is that the values obtained with analysis represent the overall *shape* of the spectrum, not only individual peaks. As inaccurate as this may sound, even 10-15 values obtained from the complex MFCC algorithm[113] are enough to give a rough idea of the shape of the spectrum without the problem of matching timbres of different pitched notes with different values. In the Score-Tool project, I call the set of MFCC values an *MFCC vector*. The MFCC algorithm is explained later, but here I can say that comparing MFCC vectors is analogous to comparing the auditory color sensations of a timbres.

The MFCC vector can be calculated from any timbre, single source, or compound, making it a sufficient tool to determine the likeness of an orchestration sound to a freely chosen target. The results should be interpreted as if the target MFCC vector resembles the orchestration MFCC vector, and if the target spectral centroid is low, then the target will likely blend with the orchestration. In comparing these two values, the preference is for the spectral centroid, because even the target with a non-matching MFCC vector can blend into the orchestration if its spectral centroid is low, as would be the case of a horn as the target playing with full strings. A high spectral centroid, however, can stand out despite the timbre similarity, as would be the

---

[113] The algorithm is explained in Part II, section 4.6.

case with a high solo violin playing with string orchestra. The MFCC vector is further discussed in the data visualization chapter.

To determine the homogeneity of the orchestration, meaning here the likeness of the timbre of individual instruments in the orchestration chord, individual sounds must be compared to one another. In this project, I compare the MFCC vectors using statistical methods. Comparison is done by calculating a Euclidian distance of vector elements 2-7 (The first element of MFCC is linked to amplitude more than timbre, and the elements after 7th one tend to vary quite little from timbre to another) of MFCC vectors of instrument timbres of orchestration. After getting distance values, my idea is to calculate the *coefficient of variation*, a method borrowed from statistical analysis, for the orchestration timbre. The coefficient of variation is further discussed in the mathematical section (see Part II, section 4.9). The idea of the coefficient of variation is to measure the dispersion of a probability distribution. This statistical tool is seldom applied to sound properties, but here it will give value to the dispersion of timbres in the orchestration. A higher value means greater heterogeneity. Low values mean low dispersion, meaning that the orchestration timbre is homogenous, thus giving user “permission” to use the MFCC comparison to test the target against the orchestration timbre. High value mean that it is unlikely that the target timbre would “stick out”, because the orchestration timbre itself is already heterogenous.

There is one caveat, which is a stumbling block for this statistical formula: If the orchestration is overly homogeneous, the coefficient of variation will be very large owing to the nature of its formula. In the formula, the mean of the samples goes in the denominator, causing the value to jump sky high when the mean dips near zero. This is the known limitation of the coefficient of variation, especially if the values used do not originate from a ratio scale. The MFCC values certainly are not a ratio scale, since it would be impossible to tell if one value has “twice as much color” compared to another. The usability of this method in orchestration is discussed later in this chapter where the formula is tested in practice.

In conclusion, based on orchestration guides, recent research on the blending of musical sounds, and my personal experiences, I hypothesize that the blendability of the target with the orchestration can be estimated by two methods in order of importance:

1. Calculating the spectral centroid of the target timbre
2. Comparing the target MFCC vector to the MFCC vector of the homogeneous orchestration.

In addition, there is a hypothesis that blending, or *cohesive sound* in the article, can be achieved while at the same time have distinguished individual parts, depending on the orchestration choices.[114]

8 PREVIOUS RESEARCH ON MASKING IN THE CONTEXT OF MUSICAL SOUNDS

The masking of musical sounds is a less well-researched topic than blending. In the present chapter, I present some of the most notable research in this field.

In 1991, Sandell stated that, in orchestration, cases of tone-masking-tone had been investigated very little, and added, “The results of such a study, even if desirable, would only remotely be

---

[114] Goodchild and McAdams 2018, p. 510.

applicable to any practical aspect of orchestration."[115] The tone-masking-tone cases are controversial, even in psychoacoustic-related papers, because of the complex phenomenon of tone interference with roughness, beating, and masking combined.

In a very early study in 1941, Pepinsky tested a simple orchestration of a B-flat major triad for five brass instruments, in which he calculated the spectra of all the instruments and determined the overlapping parts that created the masking effect. This study was an outstanding effort in the pre-computer era and in the context of turbulent times. The notable find of Pepinsky's study is that sometimes masking can be a desired effect in orchestration. For example, Pepinsky tested a chord otherwise played *piano* but raising the levels of trombone and horn to *mezzo forte*. This, according to the author, resulted in masking the overtones that produce a sensation of roughness.[116]

The masking effect has also been claimed to be responsible for the superiority of high voices in orchestral scores in general. Even in classical-era concertos, the solo instrument was usually placed at the top of the orchestra's frequency spectrum. It has also been speculated that hidden masking patterns based on pitch and register is something which composers may have implicitly known for ages.[117] In 2014, Song Hui Chon and David Huron carried out a synthesized test in which simultaneously playing instruments were placed in different orders in register, and the predominant instrument was to be identified. In three out of four cases the predominant instrument was the one playing the highest in register.[118]

There are also noise components in orchestration sound, which can be used as a basis for noise-masking-tone calculations. It has been suggested, for example, that the noise components, especially in woodwind instruments, create the softening sensation in the timbre, which is a direct effect of auditory masking.[119] The noise components still represent just a fraction of the orchestral sound world that interests me, so I do not include special cases of noise study in this project. Including noise components could be another interesting topic for future work in the Score-Tool App.

Masking in orchestration is addressed to some extent by Reuter in his musicological studies discriminating among orchestral instruments. According to Reuter, masking does not necessarily make the instrument inaudible. Furthermore, in his listening experiments, he found that micro-modulations, i.e., tiny changes in the pitch or timbre of the instrument, has a profound impact on how listeners discriminate between instruments.[120] The concept of micro-modulation is beyond the scope of my project, because my purpose is to find direct and objective data in orchestral scores to assist music professionals. Tiny changes in timbre are related to the performance of the score, although Reuter's ideas could be implemented in future versions of the Score-Tool App.

---

[115] Sandell 1991, p. 61.

[116] Pepinsky 1941.

[117] Chon and Huron 2014.

[118] Chon and Huron 2014.

[119] Sandell 1991, p. 60.

[120] Reuter 2004.

Another of Reuter's hypotheses is that the formant areas of instruments have a big role in identifying a particular instrument from the mass. The concept of formant is borrowed from linguistics, where formants define the vowel sound regardless of pitch. Orchestral instruments have also been classified according to matching vocal sounds in some studies, but the idea of instruments having formants is not acknowledged by all researchers.[121] In Reuter's case, the formant areas are the key to the idiomatic sound of each instrument. Masking the formant makes the instrument unrecognizable [122] For example, a noise-masked rendition of a synthesized oboe playing melodic figurations typical of a flute misled participants into identifying the sound as a flute.[123]

Reuter provided a Flash application for acquiring the formant data for the most common orchestral instruments. Flash is unfortunately an obsolete technique, as it was found to have irreparable security risks, and its technical support was dropped already a decade ago.

The idea of weighting the importance of formant areas in orchestration is a possible application for my project, but the data of suggested formant areas are not consistent from one author to another. Some claim that horns have one formant; others claim there are two. Moreover, the center frequencies and bandwidths of suggested formants vary considerably. This is most likely the result of analyzing different instruments played with different techniques, but even so, the frequency area which makes the horn sound like a horn is not unanimously agreed upon. The idea of formant areas would be a good place for further development of the Score-Tool App in the future.

9 PREVIOUS RESEARCH ON COMPUTER-AIDED ORCHESTRATION

The Score-Tool project can also be seen as an application for computer-aided orchestration. The difference is that the Score-Tool App is an analytical tool, not a creation tool. However, the methods for analytical and creative purposes have common ground, and therefore I want to shed light on previous research regarding creative computer applications.

The simplest form of computer-aided orchestration is playing the score in a notation program with either synthesized or sample sounds in the computer memory. This form of "help" is used by many composers today. The most common notation programs come with this feature, and efforts have been made by software companies to make this synthesized playing sound acceptable. As many composers have probably noticed, this orchestration help gives a false image of the orchestration's balance and power; weak instruments sound loud through the texture, and loud instruments blend into the mass nicely. In my opinion, the orchestration synthesis played by notation programs cannot be used as a reference point for real performance.

At the dawn of the millennium, some advanced algorithmic compositional tools were introduced that could be used to some extent for orchestration problems. These programs

---

[121] Lembke 2015, p. 29 implies that mainly orchestral wind instruments have formants.

[122] Reuter 2014; 1996.

[123] Lembke 2015, p. 9.

include Max/MSP, Open Music[124] developed at IRCAM, and PWGL[125] developed at the Sibelius Academy by Mikael Laurson. All these programs have the possibility to load a score created with a notation program, calculate adjustments, and output a score, which can be imported back into the notation program. Open Music and PWGL require some programming skills in the relatively rare programming language Lisp. These programs have been used to help with orchestration, for example, by connecting them to audio analysis modules. Max/MSP has a fairly recent library called Bach, which allows manipulation of musical scores as well as orchestration in real time. There is also an implementation of Terhardt's virtual pitch algorithm in Max/MSP, which includes the formula for calculating the mutual masking of frequency components.[126]

In a study in 2010 by Carpentier et al., three different approaches are mentioned in which a computer program was used as an aid for orchestration.[127] The first approach, by Rose and Hetrik, is an educational tool for orchestration analysis or the proposition of new orchestration for a given target sound. At the moment of this writing, that particular program cannot be found anywhere; only an article describing its features is found. It is unclear whether the intention was to make a commercial product or not. The basis of the program is the sound spectrum of combined orchestral instruments' timbre describing the orchestration. The similarity of sounds is searched by matching the spectra using the Euclidian distance method. The analysis of the orchestration is also based on each instrument's spectrum.

The second approach is a Lisp program created by David Psenicka and called SPORCH, which stands for SPectral ORCHestration. This tool does not provide analytical features, but concentrates on giving solutions to orchestrate the user-inputted sound. The search engine uses peaks instead of the whole spectrum. The search algorithm stops on the first solution, which may not be the optimal one.[128] SPORCH is still available as an open-source software and is provided as a Lisp source code. At the time of this writing, instructions for compiling the program were only for Mac OSX and require a Lisp compiler installed in the system.

The third approach is a system suggested by T.A. Hummel for orchestrating a human voice-sounding timbre with orchestral instruments. Hummel does not provide a program and does not provide any analytical method. Hummel's approach matches the spectral envelope instead of the spectrum. The idea is first to find the closest match for the target, then subtract the match from the target and continue searching with the residue of the spectrum. The method assumes that the target sound could be imitated by one strong instrument, i.e., the closest match of orchestral instruments that sound like the target, and fill in the "gaps" of the target spectrum with additional low dynamic-playing instruments.

---

[124] Assayag et al. 1999.

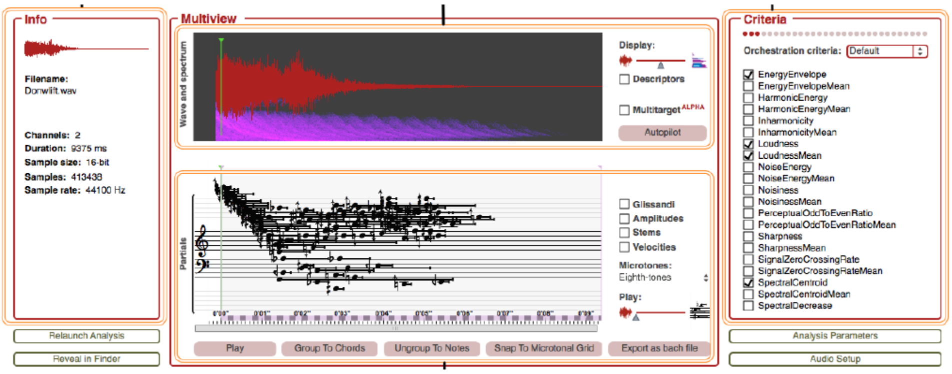
[125] Laurson, Kuuskankare, and Norilo 2009.

[126] Todoroff 1995.

[127] Carpentier et al. 2010, p. 48.

[128] Carpentier et al. 2010, p. 49.

Aside from these, the most notable orchestration program on the market has been the orchestration environment developed at IRCAM called Orchidée, which is also centered on finding a suitable orchestration for user-inputted sound. Orchidée itself is a collection of algorithms with a connection to an instrument database, allowing orchestration problems related to the realization of the desired sound to be solved. The algorithms include ideas inherited from *Patchwork*, the algorithmic composition tool developed a decade earlier by Laurson.[129] Algorithms allow, for example, limiting the search to fit the instrument combination at hand. In Orchidée, orchestration is formalized as a multi-objective combinatorial problem,[130] which means that the different aspects of timbre are used to calculate the results, and the outcome is a set of variants instead of only one solution. After Orchidée, a user interface was developed to interact with the algorithms. For a long time, the software package was sold on the IRCAM website as a commercial package, but in 2018 the software was made freely available, although the source code remained closed. Currently, the software is distributed with the name Orch-idea[131] as a Max/MSP package intended for the Mac OSX operating system.


Filename: Donwlift.wav
Channels: 2
Duration: 9375 ms
Sample size: 16-bit
Samples: 413438
Sample rate: 44100 Hz

Orchestration criteria: Default
☑ EnergyEnvelope
☐ EnergyEnvelopeMean
☐ HarmonicEnergy
☐ HarmonicEnergyMean
☐ Inharmonicity
☐ InharmonicityMean
☑ Loudness
☑ LoudnessMean
☐ NoiseEnergy
☐ NoiseEnergyMean
☐ Noisiness
☐ NoisinessMean
☐ PerceptualOddToEvenRatio
☐ PerceptualOddToEvenRatioMean
☐ Sharpness
☐ SharpnessMean
☐ SignalZeroCrossingRate
☐ SignalZeroCrossingRateMean
☑ SpectralCentroid
☐ SpectralCentroidMean
☐ SpectralDecrease

☐ Descriptors
☐ Multitarget ALPHA
☐ Glissandi
☐ Amplitudes
☐ Stems
☐ Velocities
Microtones: Eighth-tones

Example 15. Image of the user interface of the IRCAM Orchids (pre-Orch-idea) application, which is based on the Orchidée orchestration environment

All the orchestration algorithms intended to find the instrumentation for user-inputted sound could be used for my purposes, that is, to find a way to make a desired instrument audible. For example, in a piece for soloist and orchestra, the recorded soloist part could be fed into the Orch-idea to find the orchestration that would match the solo parts spectra and shown with desired time-intervals. This result could be then used for what to avoid in order to make the orchestration contrast with the soloist's part. In this way, at least blending with matching colors could be avoided. This method, obtaining results by negation, is time consuming and complex compared to the software aimed directly at the problem at hand. Therefore, I see that the development of Score-Tool App is necessary to answer the needs to which Orch-idea cannot give specific solutions.

The newest addition to the computer-aided orchestration field is the ACTOR-project,[132] whose name stands for Analysis, Creation, and Teaching of Orchestration. The project webpage

---

[129] Laurson 1996.

[130] Carpentier et al. 2010, p. 50.

[131] Orchidea website 2020 (orch-idea.org).

[132] ACTOR Project website 2020 (actorproject.org).

opened in 2020. The project gathers the latest information about orchestration-related research and is also funding its own research. Currently-funded projects include a study of combinations and contrasts, where a real orchestra plays various passages while at the same time is filmed with an acoustic camera to measure orchestral balance and sound distribution. The acoustic camera is a device that detects the power output of individual sound sources, for example, in the concert hall. This kind of cutting-edge research may provide new tools that could be implemented in computer programs to help composers create previously unheard orchestral effects.

Another Actor-funded project related to my interests is the OrchView software, which at the time of this writing is under development. OrchView allows users to annotate scores according to their perception of music. The annotations are performed by analysts listening to specific performances and looking at scores. The annotated scores are gathered into a database, which is called ORCHard, another project in the Actor community. The ORCHard project currently has a very large database of excerpts from scores of well-known orchestral works with tagged features, such as blending chords, contrasting chords, background texture, and so on. All excerpts also have listening examples from historical recordings, produced in various halls without touch-up of the dynamics. The database is searchable and is maintained by an orchestration researcher, Stephen McAdams. The database is semi-open; it is accessible to anyone with a user account, and the user account can be requested from the database webpage. The database could be used to determine the audibility of an instrument by searching the equivalent orchestration with appropriate tags and comparing that to your own. This technique is similar to that used in orchestration handbooks. From my perspective, the database can be used as test case, for example, to test passages tagged as blending chords and to try and determine with the Score-Tool App if blending is suggested in these same passages.

10 THE EFFECT OF THE CONCERT HALL ON THE AUDIBILITY OF AN INSTRUMENT

As a composer, when I write music for orchestra, I seldom think of the exact hall in which the work will be performed. Whenever I do, my thoughts are not so much on audibility, but on reverberation. If the first performance is scheduled in a hall with a long reverberation time, such as a church, I think twice about the rhythmic motives in the piece and whether they will sound blurred or not. Usually, the difference between halls in the performance is not as dramatic as my thoughts about them while I am composing.

10.1 HALL REVERBERATION

Blurred sound has an effect on perception that is similar to blurred images: the sharper image is perceived more clearly than the blurred one. Therefore, the question is, if the blurriness of sound has an effect on audibility, then is it true that the clearer the sound, the easier the discrimination? Intuitively, one might assume that a weak-sounding instrument played in the middle of the orchestra is heard better with less reverberation. In my experience, this is indeed the case most of the time. Also in general, the original idea I had in composing a new work matches the performance better in a dry hall than in a reverberant hall.

Increased reverberation adds to the perceived loudness of music because the reverberation piles on sound power that is greater than the reverberation time. The objective measure of reverberation time is defined as the time it takes for sound energy to decay by 60 dB. The average value of this decay in good concert halls is around two seconds. However, acousticians have found that the early part of the sound decay seems to determine how an audience perceives the music.[133] This early decay time is measured as the time it takes for sound to decay by 10 dB. For the average reverberation time to disturb perception, the information density of the music must be high. For example, a sixteenth note in *allegro* tempo (quarter note = 120) takes 0.125 ms to perform, but the musical harmonies change at this rate only in rare cases.

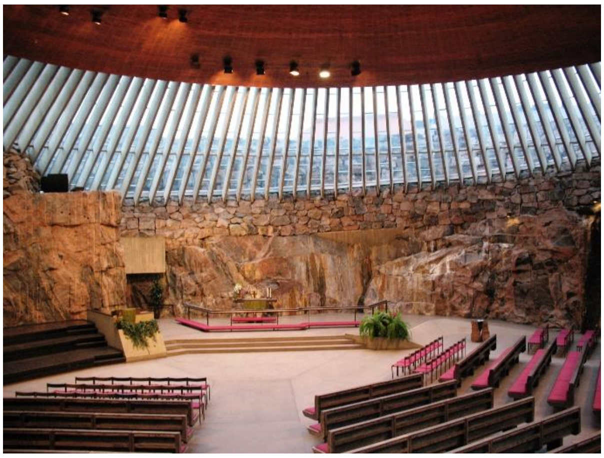
10.2 ORCHESTRA SEATING

Examples of other hall-related questions addressed to a composer in a rehearsal are whether the soloist is audible or whether the soloist should be moved to the other side of the conductor or a little bit forward (usually about half a meter). The conductor might also suggest dropping out one desk of violins so the soloist would be more audible. I have also heard both musicians and audience members say things like "In this hall, the woodwinds are heard clearly, but in that other hall, the brass dominates."

In many of these cases, the effect is very subtle and can have a bigger psychological impact than an acoustical impact on performance. Seating changes of 1-2 meters on stage would probably be unnoticeable for acoustical measuring devices placed in the audience.[134]

An exception would be acoustically projective paraboloid formations, an example of which is Temppeliaukio Church, a modern building in the center of Helsinki.[135] The church ceiling projects sound in the same way that a concave mirror projects light. My work *Sonority* was premiered in this church, and some audience members reported that the sound of the solo horn was difficult to hear, while others did not feel the same way. It might be that the ceiling projected the soloist's sounds away from some listeners, and/or projected some orchestral sounds louder to certain seats.

However, discounting this one special case, it is unlikely for a hall to boost woodwinds or make any other instrument magically audible.



Example 16. Interior of the Temppeliaukio Church, one of the few places in Helsinki where there might be significant levels of differences in loudness for the audience based on where the musicians in the orchestra are seated. The reason is that the paraboloid ceiling creates focus points for sound waves.

---

[133] Concerthalls.org website 2021.

[134] See the following section, 10.3, "Critical distance."

[135] https://temppeliaukionkirkko.fi accessed in 2021.

10.3 CRITICAL DISTANCE

In Piston's orchestration handbook dealt with above, Piston states that "placing second violins behind the first ones is one of many acoustical problems standing in the way of an exact science of orchestration."[136] The effect of audience seating on what listeners hear can be calculated. Piston's remark could be valid if we were attending a concert in the open air. Let's imagine a scene where we stand 11 meters away from an open-air stage, and the violin section plays the tone measured 80 dB one meter away from the players. At 11 meters the sound pressure would be 59.2 dB ($80 \text{ dB} - 20log_{10}11$). If the violin section moves 1 meter towards the listener, the pressure would be 60 dB, an insignificant difference of 0.8 dB in the context of orchestral music. In a hearing test, the just noticeable difference for a 60 dB noise is about 0.5 dB when the test is done in a controlled and quiet environment.[137] In other words, an audience might just barely notice the loudness difference between the first and second violins in the open air, but perhaps not in a concert hall.

Concert halls are usually reverberant places, and for those sitting in the audience the reverb will dominate the perceived sound, which means that the loudness varies very little from seat to seat. The distance after which we hear more "hall" than direct sound is called the critical distance. [138] The critical distance for omnidirectional sound is calculated with the formula $\sqrt{\frac{S*a}{16\pi(1-a)}}$, where S reflects surface area and a is an absorption coefficient. In an "average" music hall, the critical distance is just a few meters, meaning that the seating arrangements affect audibility very little. For example, Boston Symphony Hall has a critical distance of 7 meters, and almost all listeners sit beyond this distance.[139] The noticeable difference between instrumental balances in a concert hall would require at least a 3 dB drop, which means, for example, in Boston Symphony Hall sitting approximately 5 meters away from the target instrument. If an audience member's seat in Boston is, for example, 11 meters from the stage, the reverberation of the hall causes the sound to be heard louder than the direct sound for all the instruments. Based on this effect, the conclusion is that performers' seating order on stage has very little effect on audibility when the audibility is measured as the perceived loudness of the sound, except perhaps for the spatial unmasking effect.[140]

10.4 THE LISTENER'S PREFERENCE

The placement of the performer on the stage may, however, has unmeasurable consequences on music perception. There is, for example evidence that listeners cannot actually divide attention between multiple simultaneous auditory objects[141], and the auditory spatial processing engages many of the same brain regions as visual orienting[142]. Thus, seeing the musician to perform may help to make the instrument sound audible. This phenomenon is

---

[136] Piston 1955, p. 61.

[137] Fastl and Zwicker 2007, p. 182.

[138] Kuttruff 2006, p. 269.

[139] Griesinger 2011.

[140] See Part I, section 6.4, "Instrument placement in the hall and audibility" for more information.

[141] Shinn-Cunningham, Best, and Lee 2017, p. 20

[142] Shinn-Cunningham, Best, and Lee 2017, p. 26

intentionally left outside my project, and my focus is on acoustical phenomena that happens in our hearing system before neuron impulses are transmitted to the brain.

Why then do dry halls sound better to my ears? In recording practice, if a recording engineer sets the balance between direct sound and reverb, the balance is nearly always between +4 and +6 dB in favor of direct sound.[143] Musicians tend to like the same balance. Acoustician David Griesinger draws the conclusion that this might be the optimum based on human hearing, since otherwise the balance of the recordings would be mixed differently. For dry halls, the balance between direct sound and reverb is closer to this optimum than in a reverberant hall, and this might be the key to explain my preference. On the other hand, music sounds better with adequate reverberation, and in halls with reverberation, the direct sound balance is not optimum.

For my project and for the Score-Tool App, I have included the possibility of testing the effect (or lack thereof) of seating on orchestration. The App includes a model of the main concert hall in Helsinki's Music Centre based on measurements made by Tapio Lokki and Jukka Pätynen. The measurement data are used with Lokki's permission. The effect of the listening position in the hall can also be tested in the App's tutorial section, where the user can move the individual instruments and the listener in a 2D space and see the effect of the position on the sound power in the instruments. In masking calculations, the average hall reverberation is taken into account by a 0.1-second overlap of changing notes.

11 PREVIOUS RESEARCH ON AUDITORY PERCEPTION RELATED TO THE PROJECT

In this chapter, I discuss the research conducted on perceptual aspects of music which are relevant to my topic. I give a brief overview of computational instrument and timbre recognition and aspects from music psychology. I discuss perception mainly from the composer's point of view, i.e., what can be done in a musical score to enhance the audibility of the target. As I conclude at the end of the chapter, many psychological aspects that affect audibility are already included in the Western music tradition, but awareness of the mechanisms may help musicians make the best artistic choices intentionally from the audibility point of view.

In determining the audibility of an instrument, more is involved than the hearing system's physical ability to detect parts of the instrument's spectrum. The masking phenomenon is related to the mechanics of the inner ear, but blending and the whole question of audibility is also related to the cognitive perception of auditory objects and music psychology. These are topics that are hard to put into algorithms, and therefore they are only partly included in the current version of the Score-Tool App.

11.1 MIR AND TIMBRE FEATURES

Music digitalization and internet paved the way for the birth of *Music Information Retrieval* (MIR). MIR systems are used, for example, to classify large collections of recorded music on the internet so that listeners can easily find the artists they like. MIR systems are also used for analytical purposes by such experts as musicologists, music theorists, music engravers, and composers. These MIR systems have been designed with the goal of being as representationally

---

[143] Griesinger 2011.

complete as possible, especially with regard to the symbolic aspects of music.[144] In a sense, Score-Tool can be seen as a MIR system.

One branch of the MIR research is automatic instrument recognition in polyphonic music. The task is to recognize instruments from audio data, which would enable the classification of passages where specific instruments are used. Automatic instrument recognition interests me very much because the research is strongly connected to humans' perception of sound.[145] As in so many cases where computers try to mimic human perception and auditory perception is compared to the human ability to recognize sounds, the results clearly indicate an inferior performance.[146] Luckily, human perception is still the measure of all things.

The Score-Tool App estimates the audibility of an instrument by calculating values for audibility-related features of the sounding orchestration. The features related to masking, while providing a rough estimation, are easier to put into algorithms than features related to blending. Masking, especially frequency masking and its mechanisms, is a relatively well-known topic in comparison with timbre. In the Score-Tool method, the blending sensation is measured by determining the timbre distance[147], because matching timbres have been proven to blend best.[148] Blending and other timbre-related algorithms in the Score-Tool App are generalizations and simplified versions of the highly complex concept of perceptual quality that perhaps can never be mimicked by computers. One reason for this might be that masking can be seen as unidimensional quantity, like loudness and pitch, which can be measured and set to scale. Timbre, on the other hand, is a multidimensional concept, which is difficult to measure, because the perceptual mechanisms behind the sensation of timbre are yet not clear.[149]

11.2 MUSICAL COCKTAIL PARTY

Another good question is whether we, as humans, can expect to hear an orchestral timbre in an analytical way, detecting each instrument individually and also as a part of the overall timbre. A musician who listens to fully orchestrated music expects that the sounds of all the familiar woodwind, brass, string, perhaps harp, and even some percussion instruments are heard, even though some of the individual timbres might be masked by strong instruments. The expectation and also the visual feedback are strong motivations to hear even things that are not audible. As mentioned above, this human ability to focus on one sound source and suppress other competing sources is called the "cocktail party" phenomenon in speech research.[150] Human beings can also follow a conversation more successfully if the subject matter is different from that of other ongoing conversations.[151]

---

[144] Stephen Downie 2003, p. 309.

[145] Han, Kim, and Lee 2017, p. 122.

[146] Fuhrmann, Haro, and Herrera 2009.

[147] See Part II, section 8.8, "Comparison of MFCC vectors."

[148] See Part I, section 7.1,"Previous research on blending musical sounds."

[149] Fuhrmann 2012.

[150] Middlebrooks and Simon 2017.

[151] Moore 2012, p. 311

In music, the cocktail party phenomenon is different because the sources are not competing, but rather trying to create something that pleases an audience. Individual instruments can still pop up from the overall timbre and blend back in again, depending on the orchestration, our focus, and our interests. In a sense, nearly all Western music forms a cocktail party phenomenon, because very few works are composed just for one voice. The majority of chamber and orchestral works have multiple simultaneous-sounding timbres, either homogeneous timbres as in a solo piano piece or highly heterogeneous timbres as in orchestral music.

Music with multiple concurrent-sounding timbres must be our preference as humans. Otherwise, we would listen to something else. But our perception of multi-timbral music may not be as multi-timbral as the music itself. There has been some research on how many voices and instruments a human can detect in music only by listening, i.e., without seeing the players. The results are somewhat surprising, but they do give an indication that perhaps even a musician's auditory perception is not as good as one might think.

11.3 HOW MANY CONCURRENT TIMBRES CAN WE HEAR?

David Huron conducted a test with relatively homogeneous timbres, using Bach's Fugue in E-flat major, BWV 552, for organ, a work that includes passages with 1-part writing and 5-part voice writing and combinations in between. The test subjects' task was to count the number of voices they heard in specific recorded excerpts of the fugue. The subjects included both professional musicians and non-musicians. As one might guess, the musicians scored higher points in counting the voices when the number was small. One subject, a musician, stated that when the number of voices increased, he was forced to "gauge" the number by comparing the current textural density with previous textures and estimate the number of voices.[152] The overall result of the test was that the accuracy of identifying the number of simultaneous voices drops significantly at the point where a three-voice texture is augmented to four voices.[153] The drop happened both for musicians and non-musicians, indicating that three-to-four voices might be the limit for an average person. The identification depends also on source; a human voice is easier to recognize than an instrumental sound.[154]

A later test was conducted with non-homogeneous timbres by Stöter et al. (2013). The research question once again was "How many instruments can humans estimate correctly by listening?" The experiment involved 62 participants, half of whom regularly played a musical instrument. The instruments in the test were the most familiar woodwinds, brass, and string instruments in a symphony orchestra, as well as electric guitar and electric bass.[155] Participants were asked, "How many different instruments do you hear?" As in Huron's experiment, the musicians identified more instruments and performed overall about 20% better throughout the test.[156] Surprisingly, the inhomogeneity of timbres did not provide a significant advantage, even to

---

[152] Huron 1989, p. 377.

[153] Huron 1989, p. 377.

[154] Bigand et al. 2011

[155] Stöter et al. 2013.

[156] Stöter et al. 2013.

musicians, in detecting instruments. The experiment showed an assumed upper limit for items with more than three instruments.[157]

Special cases have also been reported of musicians with super-analytical ears scoring up to 90% accuracy in a nine-instrument test, in one instance, and in another, 46% accuracy in recognizing 14 out of 27 instruments.[158] However, these are special cases done with highly focused subjects in an anechoic chamber. In a normal concert hall environment with an average listener, I believe the limit of three to four sounds is reasonable.

If it is true that in symphonic music only three to four concurrent instrumental sounds can be perceived at any given moment, then what happens to the rest of the timbres? They are definitely not ignored but rather perceived as fused form. In music psychology, this fused form is called a stream. A stream is a psychological organization that mentally represents something such as a sequence and displays a certain internal consistency, or continuity, which allows that sequence to be interpreted as a whole.[159]

11.4 Musical stream

In timbre-counting experiments like those described above, the recognized instruments most likely formed an individual *stream* for the listener. While many different instruments were detected, the subject probably counted them in sequence, i.e., not simultaneously. It appears that complex sound is analyzed into streams, and we attend primarily to one stream at a time. This attended stream then stands out perceptually, while other simultaneous sound events are less prominent. We can, of course, switch our attention from one conversation to another or from one melodic line to another, and we may have some awareness of the other voices, but it appears that one stream at a time is selected for a complete conscious analysis.[160]

The auditory stream may correspond to a single acoustic source, but it does not necessarily have to.[161] The mechanisms that select the stream in musical sound constitute ongoing research in the field of music psychology. According to Moore, the timbral features that affect the stream formation are:[162]

1. Similarity: the timbres have something in common, such as pitch.
2. Good continuation: the timbre does not alter too much in time.
3. A common fate: the timbre components undergo similar transformations, for example, they start together.
4. Belonging: a single component of sound can only be assigned to one source at a time.
5. Closure: the timbre may be masked for a while, but has a good continuation compared to the situation before the masking.
6. Attention: where we switch our focus.

---

[157] Stöter et al. 2013.

[158] Fuhrmann 2012, p. 50.

[159] McAdams and Bregman 1979, p. 26.

[160] Moore 2012, p. 309.

[161] Fuhrmann 2012, p. 29.

[162] Moore 2012, p. 304–9.

For orchestration, similarity and common fate are perhaps the most significant aspects that I as a composer use in my artistic work. For example, the common technique for writing a strong melody line in the middle register is to use unison between cello and horn. The common pitch and timbral similarities enable the similarity aspect, and simultaneous note changes enable the common fate aspect.

In composition, other aspects can happen unintentionally, for example, a wide melodic leap can break the good continuation of the stream, and the timbre may unintentionally blend into orchestration enabling the belonging aspect. Also, in a concert I cannot control the audience's focus. Someone might, for example, focus on a handsome man in the viola section and try to hear what he is playing.

12 TIMBRE AND STREAM CREATION

For the Score-Tool project, it is useful to determine the key factors affecting the creation of a stream. Differences in timbre enable us to distinguish between the same note played on, say, a piano, a violin, or a flute.[163] This applies when the instruments are very different in character. In one listening test when substantial differences in timbre were introduced, listeners tended to use these differences as cues for streaming.[164]

Timbre differences, however, are not as easy a task for computers as one might think. For example, the timbral difference between the spoken words "Bass" (the musical instrument) and "pace" (a step) is minimal and perhaps would require a highly trained machine-learning model for a computer to differentiate between the two, yet native English speakers would probably hear the difference immediately. Similarly, in an orchestra there are instruments with almost matching timbres, like the English horn and the treble oboe, the violin and the viola, and perhaps even the bassoon in a high register and the horn. Comparing the spectra of these instruments' sounds might not help to distinguish them, at least without knowing exactly which parts of the spectrum to compare.

There are numerous spectral and other features that can be extracted from the digital audio. For example, the MPEG-7 standard describes a group of features for audio classification purposes, which include such qualities as harmonicity, spectrum spread, spectral deviation, spectral variation, spectral roll-off, and others.[165] These features provide information about audio perception, but in the MPEG-7 standard they are not weighted for importance. For the Score-Tool App, I wanted to include features that are significant for timbre similarity, which is one important factor in determining blend. In this version of the App, I decided to include only time-invariant features, and therefore the change rate in an instrument's spectrum over time, for example, is not counted.

In 2009, Fuhrmann et al. tested the importance of various spectral descriptors for perception. They tested the best descriptors for pitched and percussive instrument sounds and found that the energy on Bark bands, i.e., the critical band energy, was the single most important feature for instrument recognition.[166] The second best descriptor was MFCC, and the third was spectral

---

[163] Moore 2012, p. 258.

[164] Deutsch 2012, p. 219.

[165] Kim et al. 2004.

[166] Fuhrmann, Haro, and Herrera 2009.

variance over time. In fact, already in the late 1960s Plomp and his colleagues (Plomp et al., 1967; Pols et al., 1969) showed that the perceptual differences between different sounds, such as vowel sounds in speech or steady tones produced by musical instruments, were closely related to the differences in the spectra of sounds when the spectra were specified as the levels in eighteen 1/3-octave frequency bands. A bandwidth of 1/3 octave is slightly greater than the critical bandwidth over most of the audible frequency range. Thus, timbre is related to the relative level produced by a sound in each critical band.[167]

In other studies, the importance of the spectral centroid has been emphasized. For example, Hajda et al. (1997) reviewed many timbre studies and determined that the spectral centroid was the most salient acoustic measure for timbre perception in continuous sounds.[168] Tardieu and McAdams also state outright that similar spectral centroid timbres blend best.[169]

Therefore, I rely on the three features I decided to include in the Score-Tool App: the masking curve, MFCC, and the spectral centroid. These features make an efficient computational model of timbre for orchestration purposes. For example, according to Fuhrmann et al. (2009), masking and MFCC together accounted for nearly 80% of the importance for instrument recognition.[170]

Although I decided to concentrate on time-invariant features, nevertheless the *attack time,* i.e., how the instrument sound evolves over time, must be taken into account. Attack time was found to be influential in two ways: tones with rapid attacks were segregated from each other more strongly, as were tones with contrasting attack times.[171] Therefore, instruments such as percussion and the piano have an advantage in being heard through the orchestration. This partly explains the popularity of piano concertos in the orchestral repertoire, although not the popularity of violin concertos.

12.1 UNCONSCIOUS HABITS THAT MAKE INSTRUMENTS AUDIBLE

The timbre features I have discussed so far are aspects of the acoustical properties of instruments. In the Score-Tool project, I concentrate on timbre and masking, which are also within the scope of the project, but I am aware that these are not the only features for determining target audibility. Next, I will discuss other aspects related to instrument audibility from the point of view of perception, partly based on my own experience. These aspects, which can be considered already in the compositional phase, include asynchronic playing, the use of vibrato, and sound localization.

In music where simultaneous timbres have almost matching timbres, such as in string quartets, onset synchrony plays an especially important role. This is noted already in the polyphonic music of Bach, where the composer distinguishes different voices by avoiding synchronous attacks.[172] The importance of asynchronicity to make the instrument audible may even be understood subconsciously by musicians. Rasch (1978) pointed out that, in ensemble music,

---

[167] Moore 2012, p. 258 (citing Plomp).

[168] Chiasson et al. 2017.

[169] Tardieu and McAdams 2012.

[170] Fuhrmann, Haro, and Herrera 2009.

[171] Deutsch 2012, p. 201.

[172] Huron 1993.

different musicians do not play exactly in synchrony, even when the score indicates that they should. Rasch also showed that in asynchronies up to 30 ms, the notes sound as though they start synchronously.[173] Although 30 ms is a very narrow time-window, corresponding to a sixty-fourth note in tempo 120, it is enough to make different onsets stand out for the auditory perception mechanism. In my view, fast notes are perhaps not practical for the purpose of achieving better audibility of the target. I happily notate orchestral chords with synchronous onsets in the score and let the musicians work their magic for onset asynchrony.

Another interesting subconscious habit that ensures instruments' audibility is related to the mechanism of the "common fate" stream formation. According to music psychology studies, it is possible to enhance the detection of a tone with frequency modulation. For example, Rasch found that the modulation could reduce the threshold for detecting a target tone by 17 dB.[174] The most interesting thing is that this modulation was similar to the vibrato that often occurs in musical tones. It may thus be that even without altering the dynamics, playing the target instrument with heavy vibrato while having the rest of the orchestra play non-vibrato could improve audibility significantly. The vibrato affects all overtones of the instruments simultaneously, and thus applies the "common fate" mechanism to segregate the instrument into its own stream. In cases where the Score-Tool App warns of an audibility issue, the vibrato markings can be applied, if possible, with dynamic alterations to ensure the target audibility.

The effect of the orchestra seating arrangement was discussed before from the perspective of instrument loudness, but the location of the sound also has importance for auditory stream considerations. In one test where the task of the subjects was to focus attention on one specific part in a three-part composition, the error rate in detecting the target was extremely high when the three parts emanated from the same spatial location.[175] In the test, the same spatial location was created computationally by computer, since it would be impossible to place three musicians in the exact same location. However, this indicates the importance of spreading out the instruments from left to right on stage to improve the stream creation for the listener.

Deutsch also introduced an interesting hypothesis, namely, that we perceive high-register instruments better with the right ear and low-register instruments better with the left ear. Deutsch pointed out that, from the musician's perspective, the flutes are to the right of the oboes, and the clarinets to the right of the bassoons. It is interesting that the same principle tends to hold for other musical ensembles as well. We may speculate that this type of spatial disposition has evolved through trial and error because it is conducive to optimal performance.[176]

This seating arrangement serves only the musicians, because the audience, of course, has the mirror image. In particular, instruments with low registers that are to the audience's right should be less well perceived and localized.[177] This raises the question of whether the best seats in the audience, at least from the audibility point of view, are located behind the orchestra, which is possible in many modern halls around the world. Those sitting behind, however, lose

---

[173] Moore 2012, p. 293.

[174] Moore 2012, p. 295.

[175] Deutsch 2012, p. 219.

[176] Deutsch 2012, p. 223.

[177] Deutsch 2012, p. 223.

the advantage that directionality of the instrument sound makes for seats facing the orchestra.[178] Deutsch also pointed out that it is unclear how this problem can be solved to produce an optimal seating arrangement for both performers and audience.[179]

12.2 CONCLUSION OF THE PERCEPTION CHAPTER

To conclude this chapter about perception, the many features that affect an instrument's audibility happen either subconsciously in performance or are at least partly considered in concert hall design. Knowing the mechanisms, however, makes it possible to enhance the audibility for desired instruments with special arrangements. These arrangements might be a trade-off with fluent performance, at least in cases where the soloist is unusually placed.

In the cocktail party effect, which is a sum of many psychological phenomena, the results show that the sound pressure level of the sound of interest needs to be about 10–15 dB above the masking level determined by the masking sound. Otherwise, directional location is no longer possible.[180] The psychological effects therefore do not cancel the frequency masking, but are applied in cases where multiple concurrent timbres are actually heard in the auditory system.

Thus, the results of the Score-Tool App tell whether the psychological tricks for audibility can even be applied, because a completely masked instrumental sound cannot be part of the auditory stream. To ensure that an audible target instrument forms a stream, the following aspects can be taken into consideration, at least by the composer:

- Use a target instrument with a unique timbre.
- Use asynchronous musical texture between the target and the orchestration.
- Use a target instrument with a rapid attack.
- Avoid big leaps from one pitch to another because these break the "good continuation" aspect.
- Indicate vibrato for the target if possible.
- Avoid matching spectral centroids between the target and orchestral instruments.
- Mark the placement of the target instrument on stage so that it draws attention.
- Place a low-register target instrument on the left-hand side looking from the audience and a high-register target instrument on the right-hand side.
- If several instruments are chosen as targets, make sure they have matching timbres.

---

[178] See Part I, chapter 10.

[179] Deutsch 2012, p. 223.

[180] Blauert 1974.

Part II THE SCORE-TOOL APP

The current version of the app is in address http://score-tool.com.
The version which is mostly referred in this report is in address https://old.score-tool.com.
The source code of the app is in address https://github.com/upulkkis/Score-Tool-2.0

This Score-Tool project includes a computer app which can be used to pre-evaluate the audibility of a chosen target in a work's orchestration. The orchestration data can be fed as a score file from the most popular notation programs or by hand with the implemented note editor. The output from the Score-Tool App is a color-coded score indicating the strongest maskers at each 0.1 second time frame. Each time frame can be clicked to show the detailed graphs from the masking and sound color calculations.

The Score-Tool App is a Python app with a JavaScript frontend. The App can be installed locally on PC, Mac, and Linux systems that run a Python interpreter. The installation of Score-Tool automatically installs its dependencies, including the tools displaying the frontend in a browser. There is also an online version of the App, currently running at score-tool.com, which can be used even with mobile devices. The online version does the heavy calculations on a server, which makes the calculation times dependent on the server load.

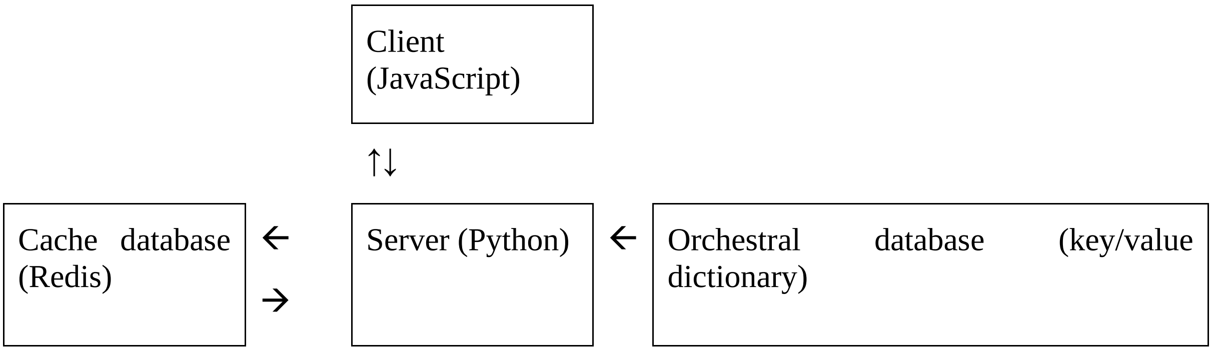
In addition, Score-Tool is available in the Apple app store, but currently only with the ability to explore one orchestration chord at a time.

1 BASIC STRUCTURE OF THE SCORE-TOOL APP

The program is divided into three parts, of which each has its own function in the process of analyzing the orchestration. The first part of the program is a script that creates the database from sound files. The database is the basis of the calculations in the second part. The second part is the server-side backend with an application programming interface (API), where the algorithms are calculated. The third part is the client-side frontend, where the user interacts with the program. The first and second parts are programmed with the Python language, and the third part with React JavaScript.

The benefit of doing the calculations at the backend is that the database of orchestral instruments does not have to be loaded into the client memory. Furthermore, the program can be used even with an entry-level mobile phone. If users want to use their own machine for calculations, the program can be installed locally, so that the client and the server are on the same machine. The user does not have to use the first part, namely, the creation of the analysis database, because the program contains a pre-created database of the most common orchestral instruments in three dynamic levels: *p*, *mf*, and *f*.

In the following diagram, the basic structure is visualized. The client interactions are sent as HTTP POST-requests and the response is returned while the client waits. The orchestral database is loaded fully into the server memory, which makes accessing the data fast. The Redis database runs on the server machine as a separate process and interacts with Redis protocol. Redis is used to speed up the program by caching frequently requested function results, so that the results can be retrieved directly from the database rather than having to be calculated again and again. The orchestral database is shared with all users, but the Redis database is individual for each user. The shared database saves server memory so that each client uses only the amount of memory that the callbacks require. This has a significant impact on memory use, because with the shared database even hundreds of clients can use the program online at the same time with just a few gigabytes of memory.



Example 17. Flowchart of the Score-Tool App functions. User (Client) exchange information with the backend server machine.

On the client side, the program can, in theory, be used with any browser. However, in practice, the user interface and some of the features are designed for Google Chromium, which is available for all the most popular operating systems: Windows, Mac OS, and Linux. Safari and Firefox browsers have some dissimilarities with Chromium with regard to the interpretation of CSS and SVG, both of which are used to render the sheet music notation on the page. The client-side JavaScript uses ReactJS library, which allows the dynamic loading of components on the page. Dynamic loading means, for example, that when a single graph is changed, the whole page is not updated, but only the graph. Dynamic loading is the key element for smooth user interactions with the web application.

On the server side, the program and the database run on a Linux virtual machine. The server is equipped with Nginx server software, which allows multiple client connections to the program at the same time. The actual Python program is served by uWSGI, through a native uWSGI protocol connection to the Nginx. Currently, the uWSGI runs in four instances on the server, each instance theoretically capable of serving thousands of connections. In practice, the simultaneous connections are limited by the bandwidth of the server internet connection and the memory of the server. The Redis cache runs as one separate process. The cached functions of each user are identified by a random string (UUID) created for each connected user separately and stored as a cookie in the client browser. If the cookie does not exist or is restricted, a new UUID is created for each session. The Redis cache is flushed once a day to prevent the cache from accumulating into a memory-hog. The Redis cache is backed up on the hard drive every 3,600 seconds. In case of power failure or server restart, all the needed services are started automatically, and the latest backup of the Redis cache is loaded into the database. If the program is installed locally, there is no need for a Redis cache, and the memory cache is used instead. This saves the user running two separate processes of the program.

The following open-source Python libraries are needed for installing the Score-Tool program locally:

- **Plotly**: For rendering the SVG data graphs.
- **Score-component:** A react component coded by me with React for rendering musical notation as SVG on a browser. The Score-component is available as open source and can be used, for example, as a basis for a Python-based notation program.
- **Music21:** for reading music xml files.

- **librosa:** For handling audio files in creating the orchestration database.
- **flask_caching:** For handling function caching in Redis.
- **Redis:** For handling the Redis protocol.
- **fuzzywuzzy:** For resolving instrument names from the score using Levenstein, an algorithm for assigning orchestral database names to the instruments.
- **Numpy:** For matrix calculations.
- **SciPy:** For machine-learning features.

To run the program and the dependencies, the Python 3 interpreter, preferably in version 3.8, must be installed along with the package manager pip. To use the frontend, a browser, preferably a Chromium-based one with access to a local host, is needed. For online use, no install is needed, just a browser.

For normal use, the online version of Score-Tool should be fine, but if a user wants to analyze a large collection of scores or customize the program to meet personal needs, local installation is necessary. The source code is hosted on Github with the name Score-Tool-2.0. For possible modifications and improvements, a separate branch can be created, which can be merged into the master branch if the code is stable.

2 THE SCORE-TOOL INSTRUMENT DATABASE

The instrument database I use is a set of audio files from recordings of the University of Iowa Musical Instrument Samples (MIS) created by Lawrence Fritt, discussed in Part II, chapter 7: "Preparing analytical data for the Score-Tool App." The entire database is about 10.6 Gb in size, which is quite large for loading fully into a server memory. For this reason, the individual sample files are loaded only when needed from a disk. However, loading files from a disk slows down the process of analyzing the score. Therefore, to speed up calculations for the end user, some calculations are made in advance. The following values are stored in a dictionary-type dataset for each sample file and fully loaded into a machine's server memory:

- Spectral peaks in dB and the corresponding Hertz values for each peak
- MFCC vector as a 13-value list
- Spectral centroid value
- Masking curve of that individual sample
- Dominant spectral value (Peaks run through Terhardt's spectral salience and spectral dominance formulas,[181] and the resulting dB values summed up using a dB sum formula)

The size of this dataset is 12 Mb, which is almost 100 times smaller than the sample library. The dataset values make it possible to construct rapidly the spectral peak constellation of any orchestration chord and calculate the masking curve without loading anything from the disk. With MFCC vectors in the memory, the comparison of timbre similarities is also rapid, as is spectral centroid comparison. The stored masking curve data speed up the calculations where one candidate is one instrument, such as a soloist target comparison against the orchestration. The latest addition to the dataset is the dominant spectral value, which gives a quick hint about which instrument in the orchestration is likely to dominate the sounding image.

---

[181] See Part I, section 5.9, "Terhardt's Virtual Pitch."

When a user calculates the masking data for the whole score, the in-memory dataset is used, but when the detailed analysis data of the orchestration chord are viewed, the full samples are used.

3 IMPLEMENTING ALGORITHMS

The algorithms can be applied only to numerical data, and therefore the orchestration data in the program appear in number matrix form. The data are acquired from the score in a similar fashion to the way the computer notation programs try to mimic the orchestra for today's composers. The difference is that in the Score-Tool App, the user cannot tamper with sound levels; the levels are fixed to mimic the acoustical characteristics of a real orchestra.

Next, I will walk through the mathematical formulas used in the Score-Tool App, which can be seen in the App's source code.

4 MATHEMATICS USED IN THE SCORE-TOOL PROJECT

4.1 DIGITAL REPRESENTATION OF SOUND WAVES

Sound pressure waves can be captured in analogue or digital media for storage, playback, and analysis. In analogue media, the shape of pressure waves can be carved, for example, on vinyl, and played back reading that carving. In digital media, the "carving" is done with numbers and a time slot is reserved to store each number. The height of the digital carving is called "bit depth" (although digital values are always either 1 or 0), and time slots can be thought of as the "graininess" of the vinyl, i.e., the smallest possible grain that is part of the carving. The most common bit depth for digital sound storage is 16, which makes it possible to represent waves that are 65536 ($=2^{16}$) units high. A common number of time slots is 44,100 per second. Time slots are called samples, so 44,100 is called a sample rate.

If an instrument sound is stored with these parameters, we get a number between 0 and 65536 every 0.00002 second. The numbers are presented in signed form, from -32768 to +32768. These numbers give the largest possible level difference between samples, the dynamic range, and the highest possible frequency that fits in the sample rate, that is, the frequency range. The dynamic range would then be 0-96.3 dB ($20*\log 2^{16} = 96.3$), which almost covers the entire dynamic range of a symphony orchestra. The frequency range would otherwise be the same as the sampling rate, but a wave has its crest and its trough, and thus takes two times the wavelength to be stored digitally. Thus, the highest possible frequency at the sample rate of 44,100 is 22.05 kHz, which is already far above any human's hearing threshold at any level.[182]

In my App, the orchestral instrument sounds are sampled at a 44.1 kHz sample rate and with a 16-bit depth.

4.2 THE FOURIER TRANSFORM

The Fourier transform decomposes a signal into standing waves of certain frequencies. In our case, the signals are digital instrument sound data. The Fourier transform is the essence of digital sound processing tools because it is the most efficient way to separate individual sine waves from a complex sounding mass. Because our time series is finite, we use the discrete

---

[182] In practice, this is a theoretical limit. To successfully analyze all sinusoidal components from signal up to 22.05 kHz, a higher sample rate than 44.1 kHz is required.

Fourier transform. Fourier’s formula is basically a correlation function of the sine and cosine waves with the stored sound:

$$X_k = \sum_{n=1}^{N} x_n e^{-\frac{i2\pi}{N}kn} = \sum_{n=1}^{N} x_n \left(\cos\left(\frac{2\pi}{N}kn\right) - \mathrm{i}\sin\frac{2\pi}{N}kn\right)$$

The spectral density $|X_k|^2$ reveals the energies of the different frequency peaks. A complicated wave consists of multiple sinusoidal components, but one must keep in mind that for each component, a full cycle must be registered in order to detect that particular frequency. Therefore, if a short sound sample is analyzed, there is a limit to the low frequencies a Fourier analysis can find. On other hand, if the analysis is made up of a long sound, we have no way to tell which frequencies are sounding at any given time.

We would like to see which frequencies are present at which times. From the spectral density $|X_k|^2$ we find the frequencies, but not at which times they occur. To obtain a more time-localized frequency information, we can smoothly cut the original signal to partly overlapping pieces, and find the spectral density for each piece. This process is called the windowed Fourier transform. The window, i.e. how to localize in time, dictates the resolution in time: and the Fourier transform of the window dictates the resolution in frequency.[183]

In my case, I am dealing with orchestral sounds that are recorded as long, steadily played tones; therefore, they can be analyzed with the Fourier transform in full resolution, i.e., without windowing. This is a huge advantage when applying mathematical algorithms to orchestration data, since even the fastest double bass tones have their base frequency peaks in place. This would not be the case if we were to analyze the same passage from an audio file played by a real orchestra.

4.3 BARK

The resolution of human hearing has been tested with a narrowband noise with a fixed center frequency, which the subject compared to a test sound with sound pressure level and center frequency equal to the reference sound.[184] When the perceived loudness change was observed, it was found that the point when the width of the narrowband noise began to affect the perceived loudness was the border of a critical band.

Using this method of measurement, the researcher Eberhard Zwicker named critical bands “Bark” bandwidths after Heinrich Barkhausen, who proposed the first subjective measure of loudness.[185] Bark bandwidths were obtained from test results of multiple subjects. Based on these results, a formula was created that gives an approximation of a critical band around the center frequency:

---

[183] The uncertainty principle in Fourier analysis, related to the Heisenberg uncertainty principle in quantum mechanics, gives a bound on simultaneous time-frequency resolution: higher accuracy in time leads to worse accuracy in frequency, and vice versa.

[184] Pulkki and Karjalainen 2015, p. 165.

[185] Pulkki and Karjalainen 2015, p. 166.

$$\Delta f_{bark} = 25 + 75(1 + 1.4\left(\frac{f_c}{1000}\right)^2)^{0.69}$$

The bandwidth is 100 at low frequencies, and it increases with frequency. In calculations, a set of 24 Bark bands is used, with border frequencies of 0, 100, 200, 300, 400, 510, 630, 770, 920, 1080, 1270, 1480, 1720, 2000, 2320, 2700, 3150, 3700, 4400, 5300, 6400, 7700, 9500, 12000, 15500, and center frequencies of 50, 150, 250, 350, 450, 570, 700, 840, 1000, 1170, 1370, 1600, 1850, 2150, 2500, 2900, 3400, 4000, 4800, 5800, 7000, 8500, 10500, 13500, respectively.

There are also other options, such as ERB representation, but I chose to use Bark, because the chosen masking formula is based on Bark bands.

4.4 MEL SCALE

A psychoacoustic test was conducted in which subjects were asked to adjust the pitch of the test tone two times higher than the reference tone. The tones were played in sequence. To a musician, it would seem obvious that two times higher than a given pitch is always an octave, but that was not the result. The subjects adjusted the test tone to approximately an octave for frequencies below 1 kHz, and increasingly large for frequencies above 1 kHz. This happened even to the point that in doubling a 2 kHz tone, the test tone was adjusted up to 15 kHz. This was explained by the low sensitivity of hearing to pitches at high frequencies.[186] The perceived pitch according to the doubling results is called the Mel pitch.

The perceived pitch curve drawn in the function of frequency can be fitted as a formula which gives a frequency to a Mel correspondence:

$$mel = 2595 * \ log_{10}(1 + \frac{f}{700})$$

This formula uses the 700 Hz frequency as the pivot point, which is also an important frequency in Terhardt’s algorithm.[187] Another coincidence is that the Mel scale and the Bark scale have something in common: 100 Mel is approximately 1 Bark.

4.5 CEPSTRUM

The name Cepstrum is a play on the word spectrum, whose first four letters are reversed. This wordplay actually tells a great deal of what cepstrum is about, namely, an inverse spectrum. The idea of cepstrum is to show the spectral envelope; in my case, this is the timbre in a convenient data form. The cepstrum analysis is mainly used in speech recognition applications, since in cepstrum it is easy to separate the glottal excitation from formants of the spoken vowels. The method can be used in my App, as the glottal excitation can be thought of as a root tone and the formants as tone color. Thus, the sound color can be inspected regardless of the notated pitch. Cepstrum is calculated as an inverse Fourier transform of a logarithmic magnitude spectrum.[188] The magnitude spectrum is an absolute value of a spectrum with

---

[186] Pulkki and Karjalainen 2015, p. 174.

[187] See Part I, section 5.10, “Spectral dominant region.”

[188] Pulkki and Karjalainen 2015, p. 56.

imaginary parts, and the logarithm represents the nature of our hearing system. The spectrum is therefore treated as if it were a sound signal:

$$cepstrum = fourier^{-1}(\log|fourier(x(t))|)$$

In my App, the cepstrum is not used directly, but rather as part of calculating the MFCC values described next.

4.6 MFCC

The abbreviation MFCC stands for Mel Filter Cepstral Coefficients. It is a method for efficiently describing the vowel formant structure in recorded speech. The calculation itself is computationally heavy, but as result gives a comparable set of numbers that describes the timbre numerically. The calculation of the MFCCs starts with an ordinary Fourier transform calculation of the signal. The sound spectrum obtained by Fourier transform cannot be simply converted to a Mel scale, but we can use a filter bank constructed according to Mel frequencies to obtain the desired result. After applying the filter bank, the first phase of the cepstral calculation is applied: a logarithm of powers at each of the Mel frequencies. It has been shown that the formant locations on this side-result are located in multiples that correlate with the increasing oscillations in frequency of a cosine wave.[189] The coefficients are obtained by applying a discrete cosine transformation to the side-result to obtain the end result. The first coefficient is usually omitted, as it mainly contains information about the sound power of the signal. The coefficients after that describe the sound color. Usually, 10-15 of the first coefficients are used, since thereafter the values are non-informational. In my App, I use 13 coefficients:

$$MFCC_i = \sum_{k=1}^{20} x_k * \cos\left(i * (k - 0.5) * \frac{\pi}{20}\right), i = 1,2,\ldots,13$$

where *i* is the coefficient number of calculated MFCC's, and *k* represents the log energy output of k:th mel-filter in x_k.[190]

4.7 THE SPECTRAL CENTROID

The spectral centroid is equivalent to the center of mass in real world physics. For a sine wave, the spectral centroid is the same as the tone itself, and for a harmonic-rich sound, the centroid is a lot higher than the root tone. In orchestration, this value gives an estimate of the brightness of the sound. Sounds with a rich spectrum, an oboe sound, for example, are generally thought to be brighter than sounds with just a few overtones, like the sound of a low flute. Brightness is not always the right attribute, because the root tone affects the brightness sensation. Therefore, it is clearer to talk about the spectral centroid. The centroid is calculated from an absolute valued Fourier transform of the sound as the weighted mean of the frequencies with the magnitudes as the weights:[191]

---

[189] Davis and Mermelstein 1980.

[190] Davis and Mermelstein 1980.

[191] Peeters 2004.

$$spectral\ centroid = \frac{\sum_{n=0}^{N-1} f(n)(x(n)}{\sum_{n=0}^{N-1} x(n)}$$

The result is a frequency which is the center of the mass of the current sound.[192] In orchestration, this center of mass gives a hint about whether the sounding image is dark or bright.

4.8 MASKING SPREADING FUNCTION

The auditory masking of a sinewave sound component does not necessarily affect only one critical band. If the sound component is loud, then the masking spreads to neighboring bands. This spreading can be expressed as a function that describes how much each neighboring critical band is masked by the single excitation. The masking spreads very little to the frequencies below the excitation, yet very much to the frequencies above. For my App, I use the spreading function formula that was created for audio coding purposes. Though the formula is fairly simple, it has proven to be functional in masking quantization noise, since audio coding glitches could irritate the majority of listeners. The spreading function for each spectral peak is calculated as follows:

$$spreading\ f(x) = \begin{cases} 17 * (dz+1) - \left(0.4 * X\left(z(j)\right) + 6\right) dB, & -3 \le dz < -1\ Bark \\ \left(0.4 * X\left(z(j)\right) + 6\right) * dz\ dB, & -1 \le dz < 0\ Bark \\ -17 * dz\ dB, & 0 \le dz < 1\ Bark \\ -(dz-1) * \left(17 - 0.15 * X\left(z(j)\right)\right) - 17\ dB, & 1 \le dz < 8\ Bark \end{cases}$$

where z(i) is the index of the spectral line (taken from a table of MPEG coding in ISO/IEC 1996, p. 89), z(j) is the masker, and dz is distance in Bark dz=z(i)-z(j). X(z(j)) is the sound pressure level of the j:th masking component in dB.

4.9 COEFFICIENT OF VARIATION

The coefficient of variation is usually not part of an audio analysis toolbox. When applied to a group of values, the coefficient of variation gives a value of the dispersion of a probability distribution. In statistical analysis, the coefficient of variation is used for comparison between data sets with different units or different means, since in those cases it gives better results than the standard deviation. The formula for the coefficient of variation (C_v) is quite simple:

$$c_v = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2}}{\frac{1}{N}\sum_{i=1}^{N} x_i} = \frac{standard\ deviation}{mean}$$

When applied to a group of MFCC vectors in an instrument sound sample, as in my App, the coefficient of variation tells us the homogeneity of the orchestration. If the orchestration consists of, say, a string quartet, the C_v is very low. If we add a trumpet to the quartet, the C_v value will rise. Adding a percussion instrument, timpani, for example, causes the C_v to be very high. As the formula shows, when the mean goes very low, the resulting value jumps sky-high,

---

[192] The centroid frequency is not related to the overtone structure in any way, i.e., it is unlikely to be the same frequency as any of the sound's overtones.

This has been found to be a problem for the general statistical use of the algorithm.[193] In the Score-Tool App, the value is presented as a percentage from 0 to 100, where a lower value means more homogeneity. The coefficient of variation is further discussed in Part II, section 8.6, “The homogeneity of the orchestration”.

4.10 TERHARDT’S SPECTRAL DOMINANT REGION FORMULA

Terhardt’s spectral formula shows the weight of individual spectral peaks in relation to a spectral dominant area. The formula is constructed according to tests of speech intelligibility.[194] The weight values, which Terhardt et al. call *spectral-pitch weight*, can be obtained from the spectral peak data of the sound with the following formula:

$$\left(1-\exp\left(\frac{-LX_{\mu}}{15\,dB}\right)\right)*1+\sqrt[-2]{0.07(\frac{f_{\mu}}{0.7\,kHz}-\frac{0.7\,kHz}{f_{\mu}})^{2}}$$

where LX_μ is the SPL excess of the µth spectral peak, and f_μ is the frequency of that peak. According to Terhardt, “it is presumed that the spectral-pitch weight describes the relative salience of competing spectral pitches.”[195] I interpret the formula to mean that these, the most salient spectral pitches, are the ones that form the essence of the sound’s timbre.

---

[193] Everitt and Skrondal 2010, p. 78.

[194] See Part I, section 5.10, “Spectral dominant region.”

[195] Terhardt, Stoll, and Seewann 1982, p. 683.

5 INTERPRETING DYNAMIC MARKINGS IN ORCHESTRAL SCORE

5.1 ORCHESTRAL INSTRUMENT DYNAMICS IN SCORE-TOOL

In this chapter, I first briefly discuss the range of dynamic markings used from the classical period to the present and the problems that may arise when different dynamic markings are used in simultaneous timbres. I also go into deeper detail on what happens to the timbre when an instrument is played at different dynamic levels. In addition, I discuss the latest research on instrumental sound directionality and how this affects the measurements of an instrument's SPL level. At the end of the chapter, I explain how and from what sources I gathered the analytical data for my Score-Tool App. I start by defining the most frequently-used dynamic markings in musical scores.

Dynamic markings, which composers use to convey a desired sound power, are not absolute values, but rather relative ones. For example, today when a composer wants *mf* from a violin, the way this desire is expressed depends on the context. This sound power, which is *mezzo-forte* in chamber music, would be *piano* in a full orchestral *tutti* chord, but perhaps *forte* in a solo piece. In other words, dynamics are like salary: a mediocre (*mezzo-forte*) salary in a middle-income country would be a poor salary (*piano*) in a rich country, but outstanding (*forte*) in a poor country.

5.2 THE DYNAMIC PALETTE

The dynamic palette used in a score varies considerably, depending on the composer, the historical time of composition, the style and mood of a piece, and the instruments and orchestra for which a work is written. In the eighteenth century, the dynamic palette was not as broad as it is today. To simplify, music was either loud or soft. The different shades of loudness and softness came later, when instruments evolved so that a wider dynamic palette became possible. For example, in Mozart's piano sonatas written for the fortepiano, an instrument capable of producing loud and soft sounds and many shades in between, only a handful of *fortissimi*, *pianissimi*, or gradational markings can be found.[196] The use of extreme markings had not yet undergone the inflation it would experience a little more than a century later. Mozart's *ff* undoubtedly meant *as loud as possible*. If you find many *ff*s and *pp*s in Mozart's score, it is probably an edition in which Mozart's dynamic markings have been adjusted, such as Grieg's edition of Mozart's sonatas.[197]

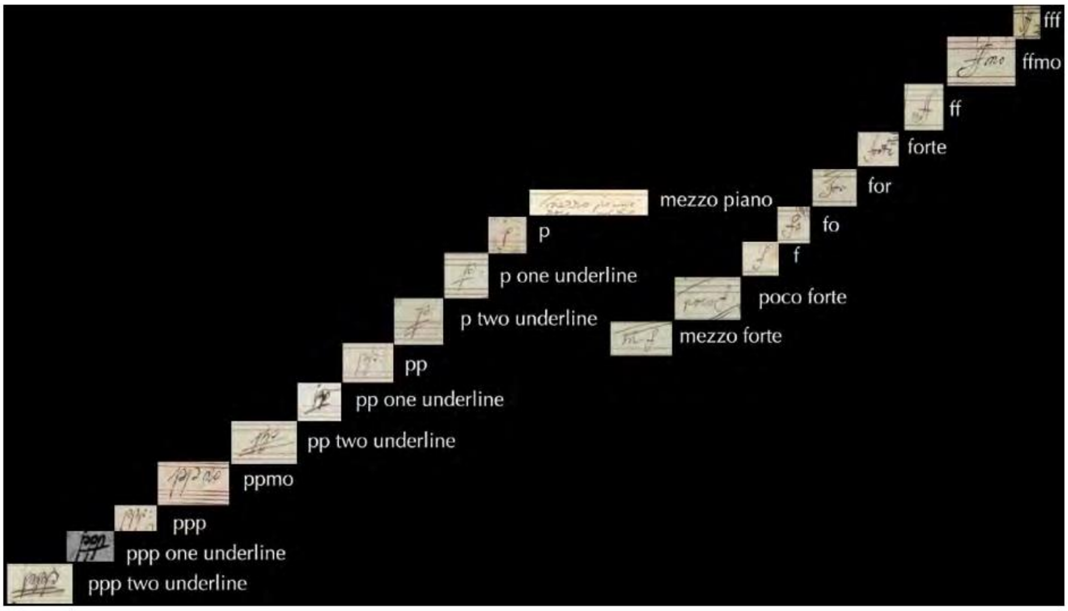
Beethoven, only fourteen years younger than Mozart, used many different dynamic markings, sometimes underlining a marking to emphasize that it was accurate. He even used an interesting variety in the middle-dynamic range, such as *poco forte* (almost loud) together with the usual *mp*s and *mf*s. Example 18, borrowed from Nicholas Kitchen,[198] shows the variety of dynamics used by Beethoven in his manuscripts, especially in his late works.

---

[196] Silverman 2020.

[197] Noh 2009.

[198] Kitchen 2017.



Example 18. Handwritten dynamic markings in Beethoven's late works. The markings are organized from soft to loud (Kitchen 2017, p. 27).

About the time Berlioz published his *Treatise on Instrumentation* in the 1840s, a greater variety of dynamic markings than before began to appear in musical scores; whether or not this was due to Berlioz is not clear. The trend escalated. The *ff* marking, as loud as possible, was not enough when a composer, for example, wanted the sound to be even louder than a colleague's *ff*, or wanted to use the increased dynamic range of newly developed versions of orchestral instruments. By 1900, it was not uncommon for composers to use as much as a 14-step dynamic palette ranging from *pppppp* to *ffffff*, found, for example, in some passages by Tchaikovsky and Puccini. These markings are, of course, absurd. One can only imagine how a crescendo from *ppppp* to *fffff* would sound.

The loud markings beyond *ff*, in my view, are unnecessary, because the possible variety of playing loud is much smaller than the variety possible in playing soft. A diminuendo from *fff* to *ff* is unimaginable, but a diminuendo from *pp* to *ppp* is not. This can also be seen in works by composers interested in timbre. Especially for impressionists such as Debussy and Ravel, the dynamic range was extended especially in the softer end, where the markings *poco pp* and *piú p* are extremely useful.

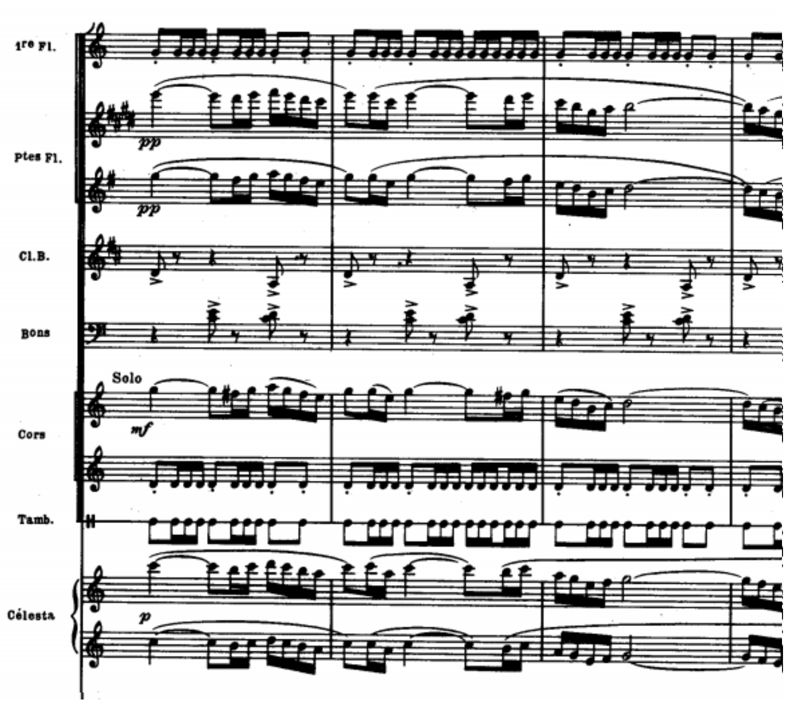
After many experiments in dynamics before and after the world wars, dynamic markings in composers' scores settled down to a reasonable level. In contemporary music today, discounting the extremes, the useful palette is perhaps from *ppp* to *fff*, resulting in an eight-step scale, which, together with careful orchestration, should be sufficient for notating the timbres in a composer's mind. Some *piú* and *poco* markings are used when necessary, depending on the context.

Most composers I know use dynamics in the context of their current piece. If the mood of the piece is soft, then a wide palette of soft dynamics is used and *vice versa*. In orchestration the almost eternal and unanswered question is this: Are the dynamics marked in relation to the overall timbre, in relation to the weakest or loudest instrument, or in absolute values? This issue is not as significant when a composer relies on the conductor's expertise in balancing the music, and thus the composer will write uniform dynamics throughout the score. It becomes significant if the composer wants to experiment with timbres involving unusual balance.

5.3 MARKING DYNAMICS FOR CONCURRENT TIMBRES

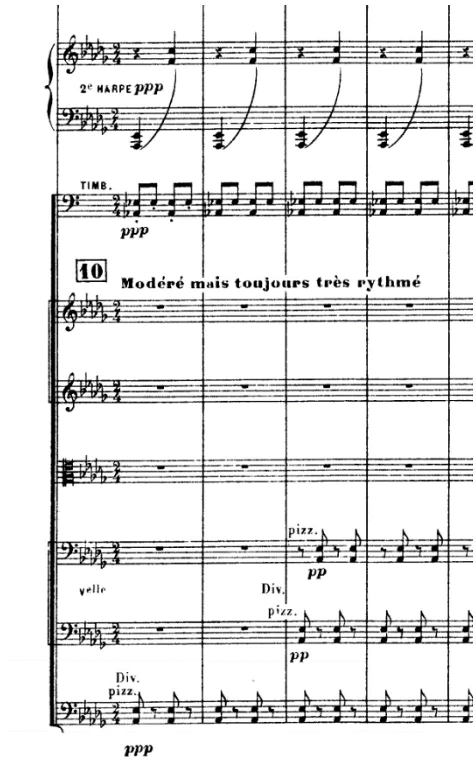
In classical and early romantic orchestral pieces, different dynamics are seldom specified for simultaneously sounding instruments. However, from the composer's perspective on the art of orchestration, asking for different dynamics for concurrent timbres is the most interesting procedure, but also the most troublesome. Different dynamic markings give a composer the opportunity to balance timbres in new ways, and that in turn opens an unlimited timbral palette compared to the somewhat limited range of uniform dynamics.

For me as a composer, dynamics are a constant headache in writing orchestral music. Often, I feel the need to treat dynamics like absolute values, because doing so gives a sense of control over the orchestration. In passages where the timbre parameter is not the most important one, I often mark double bass dynamics a notch louder than the rest of the orchestra, especially when the basses play *pizzicato*. Also, in *tutti* chords I tend to avoid writing *forte* for upper-register brass unless I really mean it. For me, a woodwind *f* equals a brass *mf*, as is the case for many other composers. The piccolo is another instrument for which I hesitate to give loud dynamics, but mostly because my ears are sensitive to high pitches.

Sometimes when I double melodies in unison or in octaves, I mark the predominant melodic line with a higher dynamic level than the doublings. This method has been used by other

Example 19. Ravel, *Boléro*, mm. 3-5 after rehearsal number 8. The composer gave three different dynamic markings for the horn melody, which is indicated as a solo and doubled by the celesta and two flutes.

composers as well. In the famous passage from Ravel’s *Bolero*, the melody is played by a horn and doubled by celesta and flutes, seen in Example 20.

All three timbres are marked with different dynamics, and at first glance the interpretation seems clear. But a closer look reveals ambiguities. The part is marked for solo horn and, since by nature the horn is the loudest instrument of the three, in performance it should be audible. Given this instrument’s nature, the *mf* of the horn, a dynamic which is just a notch below *forte*, could be thought to be quite a loud sound. The flutes, especially the higher one, are in a register where *pianissimo* is possible, but the passage needs attention from the performer to go against the most convenient dynamic in that register, which would be around *mf-f*. The celesta’s dynamic marking is almost irrelevant, because the dynamic range of the instrument is very narrow. The following questions of balance may arise: Is the *mf* of the horn to be interpreted in light of the dynamic range of the weakest instrument of the three, the celesta? Is the *p* of the celesta something that happens “automatically,” or should the celesta play *piano* in relation to its dynamic range, resulting in a sound pressure comparable to something like *ppp* on the horn? Should the flutes really play at a dynamic level a notch softer than the celesta? And does that mean that the celesta sound should be the dominant one for the listener? How would the balance change if the dynamics of all three were marked *pp*? All these questions are answered by listening to the excerpt in concert. The result sounds like a super instrument, where the fundamental tone comes from the horn with some whistle added by the flutes and a glimmer by the celesta. Perhaps asking the celesta player how to interpret the dynamic level would be a silly question. The player’s innate musicianship automatically interprets the moment in a way that “the music demands.”

Another example, from Debussy’s *Nocturnes*, appears in a passage starting at rehearsal number 10, where the cellos are marked a notch louder than the double basses, timpani, and harp.

Example 21. Debussy, Nocturnes, mm. 1-4 from the rehearsal number 10. The composer marks a different dynamic for celli (the second and third lowest staves) that double the double basses (the lowest staff) an octave higher.

The Example 21 is not as ambiguous as the Example 20 before but is still a statement from the composer. At least the double basses and timpani are by nature larger and louder than the cellos, and therefore the cellos need a louder dynamic to match the one written for the others. The problem is that by doing so in this particular passage the question is raised of whether the other passages should be interpreted as if the dynamics are absolute. For example, in the *tutti* chords in *Nocturnes* where all the instruments, including the brass, are marked with *f*, should the brass be louder? Probably not.

Using the same analogy as before, if an IT company has offices in low-income, middle-income and high-income countries, should the employees doing similar jobs in different countries have the same salary? Probably not.

Usually, in my experience, when I write different dynamics for different timbres, the conductor gets the point by reading the score, and there is no need to clarify my intentions before or during the rehearsals. A wide selection of dynamic markings is also an indication to the conductor that the dynamics are important. That is why it is more advisable to mark the balance adjustments with *piú* and *poco* rather than with a completely different dynamic. For example, in a *tutti* chord in *mf*, it would be wise to mark the brass as *poco mf* (not *mp*) and the viola as *piú mf* (not *f*). If the nature of the music is such that the dynamics are thought to be absolute throughout the piece, it can be so indicated in the score, as is the case in *Speakings* by Jonathan Harvey, where the composer states, "[Take] great care to respect the dynamics."[199] The dynamics are very important for Harvey's piece, because they are calculated with the IRCAM software Orchideé, which, according to Carpentier, was "very helpful in finding the finest balance in the instrument intensities." [200]

Generally, my view is that the dynamics in a score should be as intuitive as possible and should be reinforced by writing for the instruments in registers that result in giving the desired dynamics without an effort. In other words, one does not ask an oboist to play the lowest notes of the instrument *pp* or the bassoonist to play the highest notes *ff*. Instead, one replaces the oboe with the clarinet and the bassoon with the trombone. In orchestral rehearsals, there is no time to balance each and every chord in a piece. Therefore, the more a score balances itself, the more time it saves in rehearsals. The Score-Tool App can be used to check whether the instruments are in the right balance, especially in cases where the solution is not as obvious as in my earlier example with oboe and bassoon.

Because dynamics are such a problematic subject in dealing with orchestral scores, I decided to use only a three-step dynamic scale in the Score-Tool App. Each instrument is analyzed playing only *p, mf,* and *f*. I made this choice because otherwise, every score composed with the help of Score-Tool would require a text similar to Harvey's on the title page. There is, however, a means to check how the timbre balance changes by fine-tuning each dynamic level, i.e., adjusting *p* to correspond with everything from *ppp* to *piú p, mf* to *poco mp* to *piú mf*, and *f* to *poco f* to *fff*. By using the fine-tune slider instead of the exact dynamics, I want to emphasize once again the directive nature of the App. It does not give ready solutions but instead pushes the user in the right direction.

---

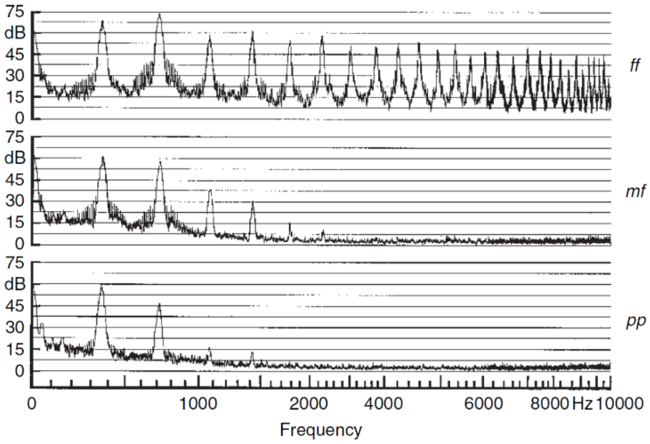
[199] Harvey 2008, p. 1

[200] Carpentier et al. 2010, p. 60.

5.4 HOW DYNAMICS AFFECT TIMBRE

If we consider a single sine wave, the sound power depends solely on the amplitude of the wave. On a frequency graph, a *p* sine sound is a short spike, and an *f* sine sound is a long spike. The overall sound power, however, depends on the power of all spectral components, so if the sound has rich timbre, i.e., has many overtone components with high amplitudes, the overall sound power goes up. Therefore, when a player increases the dynamics in playing an instrument, the amplitude of a single frequency component may not rise that much, but the overall sound power is louder because the number of spectral components rises.

This is most notable on clarinets and brass instruments, where, especially with the brass, the spectral components added by the rise of the dynamics create the distinguishable "brassy" sound in *forte* passages. This can be seen clearly on the frequency graph in Example 22, where there are just a few spectral components in the sound when the horn is playing *pp*, but over twenty components when it is playing *ff*.[201]



Example 22. From Meyer 2009, p. 35. The long-time spectrum of the horn sound played at three dynamic levels. The overtone structure is much richer when played *ff* than when played i*pp*.

As seen on the graph, the SPL level of the fundamental frequency is approximately only 5 dB louder at *ff* compared to *pp*. This affects the masking curve such that *ff* playing does not necessarily result in more masking on a single critical band, but rather results in more masking on several bands. Another study shows that the range of the level in dynamic loudness of musical instruments depends not only on the dynamic sound-pressure level range, but also is influenced by changes in the spectral envelope that accompany the dynamic gradations of sound.[202] This may explain why there might be surprises in audibility in orchestral scores; the *forte* playing can spread the masking phenomena upwards in register more than expected.

The different registers of the instrument also change the sound spectrum. For example, the lowest tones of a clarinet have a strong emphasis on odd partials, which can be tested with the Score-Tool App by choosing low tones as the target. This emphasis disappears somewhat in

---

[201] Meyer 2009, p. 35.

[202] Miśkiewicz and Rakowski 1994.

the clarinet's middle and high registers. The bassoon spectrum also varies considerably according to the register. Depending on the instrument, there are registers where it is impossible to play soft or loud, an issue which is discussed later.

Lastly, players who regulate the intensity of the sound will of course use their experience and ears to adjust the power level according to the dynamic indications in the score. The resulting SPL level depends thus on the register of the played pitch, because the equal loudness curves[203] already vary a great deal in the range of musical pitches[204]. It means that, for example, instruments with low frequency partials need a higher SPL level in order to attain loudness equal to the instruments with loud partials in sensitive frequency area. Thus, a double bass player's *pp* has significantly more decibels (SPL) than a flutist's *pp*. Also, the highest pitches of the piccolo even played *pp* will sound loud because our hearing system is very sensitive at those frequencies.

In 1991, Sandell compiled a database of orchestral instrument spectra called SHARC, which was intended to give composers and musicologists the means to investigate orchestration solely on the basis of the change in the spectrum when a pitch or the dynamics change.[205] The sound power parameter, i.e., how loud instruments are in relation to each other, was left out of the database, perhaps partly because of insufficient data, partly because the researcher thought that the spectrum change has such a huge effect on timbre that the dynamic parameter was not needed. The database in the Score-Tool App serves a somewhat similar purpose to that in the now more than 20-year-old SHARC database, but the Score-Tool App database is in extended form where more instruments, playing techniques, dynamics, and analytical data about timbre are included.

5.5 MEASURING AN INSTRUMENT'S DYNAMICS

What I have found is that there are at least three different needs for measuring the sound power of different instruments in a symphony orchestra. The first is an analytical need, i.e., curiosity on the part of either acousticians, musicologists, or composers to know how loud an orchestral *forte* is, how soft an orchestral *piano* is, which is the loudest and the softest instrument measured in dB, and so on. I also address this need with my Score-Tool project. The second need has to do with health and safety, namely, investigating the amount of sound power to which musicians are exposed while playing in an orchestra. These investigations are interesting because an instrument's sound directivity is often measured in order to know where to apply damping on the stage. The third need is technical, a need often expressed by recording engineers who want to use the right equipment to capture as much of the instrument sound as possible. Technical measurements are often made several meters away from the player, in a place where the microphone would be in a recording session. In measurements, the range of intensity is estimated for each instrument in order to choose the right microphone and set the right levels so that there will be minimum amount of noise and the sound signal will not exceed the recording equipment's dynamic limit.

Optimally, measurements of the audience's seating in a good hall for every instrument at every dynamic level would be great, but no such studies exist to my knowledge. The reason is

---

[203] See Part I, section 5.3.

[204] Equal-loudness contours are only valid for pure sine tones.

[205] Plazak, Huron, and Williams 2010.

probably that it would serve a purely analytical need, and the data would not be beneficial for either health or technical needs. The orchestra hall acousticians are, of course, extremely curious to know how the orchestral sound is perceived by the audience. However, their interest is focused more on the overall timbre rather than on individual instruments.

In practice, it is the conductor who has the final word in balancing the music. The conductor's podium has proven to be a good place to listen to the balance, at least when the orchestra uses standard seating. Pätynen and Lokki (2016) showed that the perceptible dynamic ranges in halls are more homogeneous near the orchestra. In such positions, the sound field is typically dominated by the direct sound, and the angular spread of the sound sources on stage is wider.[206] The dynamics a composer writes are thus first and foremost for the conductor, who decides whether the music they hear matches the impression the score gives. On the podium, where the orchestral timbre is less diffuse than on the audience's side, the directivity of the instrumental sound plays an important role. The balance the conductor hears might differ considerably from what the audience hears. The directionality of the sound even changes noticeably with the register of the tone played.[207] Therefore, it is extremely important for a composer to listen to the timbral balance in rehearsals and make dynamic adjustments in the score if needed.

5.6 DIRECTIONALITY OF THE INSTRUMENT SOUND

An orchestral instrument is not like a spherical buzzer that radiates soundwaves in all directions, nor is it like a loudspeaker that directs soundwaves in one specific direction. The directivity of orchestral instruments' sound, i.e., the direction in which the sound is heard at its loudest, is a complex phenomenon. This has been less thoroughly researched than instruments' sound radiation patterns, meaning which part of an instrument emanates which frequency. There are a few studies about the sound and directivity of the symphony orchestra instruments, such as Meyer's *Acoustics and the Performance of Music*,[208] but unfortunately, a detailed description of the measurement method is not available.[209]

Some instruments have a relatively simple directional pattern, such as the trumpet, which shoots at least high partials of its sound mostly where the bell points. A simplified graph in a two-dimensional plane, given in Example 23, shows how the directionality of the trumpet sound works as seen from above the player.[210]

---

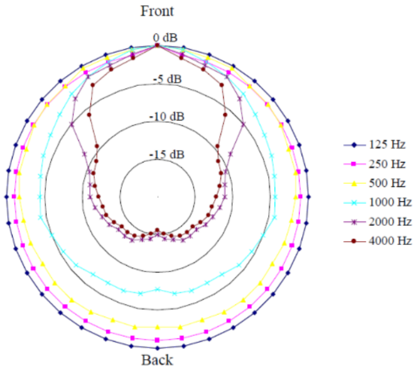
[206] Pätynen and Lokki 2016.

[207] Pätynen 2011.

[208] Meyer 2009.

[209] Pätynen and Lokki 2010.

[210] Inácio 2005.



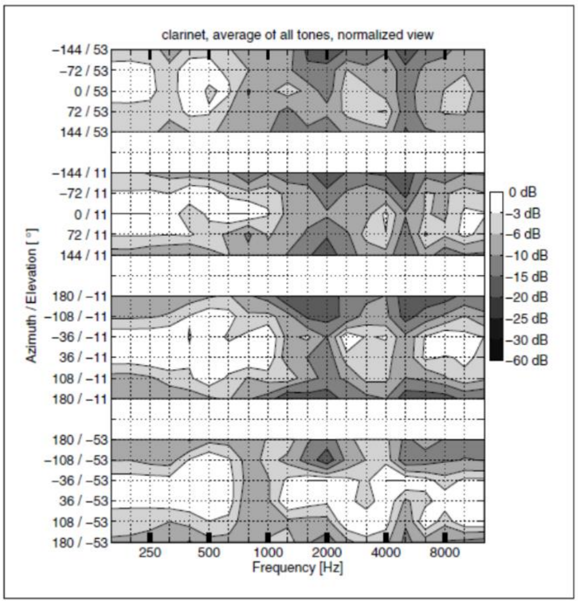
Example 23. From Inácio (2005), p. 9. Directionality of the trumpet's sound when the trumpet bell is facing front (front being the top of the figure; the player's head is in the middle of the circle). The low frequencies radiate in all directions, and frequencies above 2 kHz are highly directional. Each circle shows how much the frequency is attenuated in that direction; for example, 4000 Hz is attenuated -15 dB to the back of the player compared to 0 dB to the front.

The pattern shows how the fundamental tones radiate in all directions, but higher partials, the ones that our hearing system register the most, radiate almost solely in the direction of the bell. In other words, in an open-air performance the trumpet can be heard with added clarity where the bell points. In an orchestra hall, the reflections smooth out the directionality, but even there directionality can have consequences. For example, in a trumpet concerto where the soloist stands beside the conductor pointing the instrument at the audience, the conductor might not hear the balance as it sounds to the audience. The conductor might think that the audibility of the trumpet is poorer than it actually is and make the orchestra play softer than needed.

All instruments have unique directional patterns, which depend not only on the instrument's mechanics, but also on the playing style. For some instruments, different frequencies radiate in seemingly arbitrary directions, making it hard to predict the actual place where the instrument's sound has its nominal characteristics. For example, the sound radiation pattern of the clarinet is shown in Example 24, "photographed" (or better put "audiographed") with a microphone array placed above, below, left, and right of the instrument in an anechoic chamber.[211]

---

[211] Pätynen and Lokki 2010.



Example 24. From Pätynen and Lokki (2010), p. 146. The directionality of a clarinet sound shown in Jukka Pätynen's research. Four graphs represent a two-second sound field at different heights, marked as "elevation" in the graph. For example, if we read 500 Hz directionality on the second graph from the top, which shows readings at 11 degrees of elevation, we see that the 500 Hz frequency attenuates up to 10 dB to the side (144 degrees azimuth on the graph).

The graph shows how unevenly different frequencies emanate from the instrument's body. Especially in the 2 kHz area, up to 20 dB attenuations depend on the listener's position.

Although the full symphony orchestra creates a complicated web of frequencies directed to all imaginable directions in the 3D field, the instrument measurements and recordings can still be used to estimate orchestration balance, because the directionality of the different frequencies is smoothed by distance. There is an additional coefficient in the formula for critical distance[212] for directivity of the sound source, which makes the critical distance longer if the sound is highly directional. In a concert hall, the majority of seats are well beyond the critical distance of even the most directional instruments and frequencies. Still, there is a slight possibility that some frequencies may be more distinguishable at one seat than at another. The reason is that the ratio between direct sound and reflections change, depending on the listening position, although beyond the critical distance the timbre is dominated by reflections. Some of these reflections can be quite dominating, for example, if the horns point their bells to the ceiling in the back of the hall, and the ceiling surface is hard.

5.7 FROM DYNAMICS TO DECIBELS

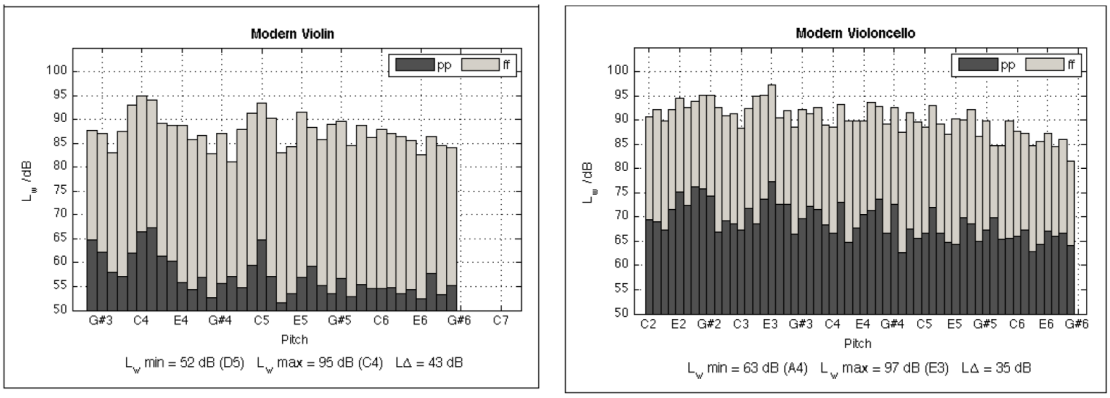
Before even going into this subject, I want to start by saying that there is no general way to convert dynamic markings to decibels. No such thing can exist. There are just too many factors contributing to the interpretation of dynamics for a single formula to show which dynamic marking corresponds to a specific dB level. Especially in the low dynamic range, the sound power that actually comes out depends on the style of the piece, the texture played, the instrument, the pitch, the mood of the piece, the mood of the player, an orchestra's unwritten

---

[212] See Part I, section 10.3, "Critical distance."

rules, the conductor's interpretation, the hall, and more. For some reason, the problem is smaller with loud dynamics, perhaps because our hearing system is more sensitive to dynamic changes in quiet sounds than to dynamic changes in loud sounds. For example, in one study the just noticeable difference was significantly smaller for noise sounds around 60 dB SPL than for noise sounds around 80 dB SPL.[213]

In a study where the aim was to investigate the dynamic levels of historical and modern instruments, musicians were asked to play scales both *pp* and *ff*. My interest is naturally in the levels of modern instruments. The resulting sound power was measured from three meters away with calibrated equipment. The resulting graphs are shown in Example 25.[214]



Example 25. From Krämer (2011), p. 21. Dynamic levels of different pitches played by a violin and a cello in an anechoic chamber, recorded with multiple microphones a few meters away from the instruments. The musicians were asked to keep each dynamic level constant. In both instruments, there is more fluctuation in interpreting different pitches at the *pp* level than in interpreting pitches at *ff*.

The first thing to notice in the graphs is that there is not much difference between the *ff* of the violin and the *ff* of the cello, but there is a significant difference between the two instruments at the *pp* level. Especially given the variance in the sound power at *pp*, it is clear that the *pp* dynamic is far from absolute. For the violin, the loudest *pp* notes around C4 are about 67 dB (SPL) loud, and the softest around D#5 are about 52 dB (SPL). There is a 15 dB variance for interpreting *pp* dynamic within the played notes for single instrument and single player. Part of this difference can be explained by the equal loudness curves, discussed above in "How dynamics affect timbre" (section 5.4). Changing from one string to another also affects the dynamics. The essential lesson may still be that even the best artists do not have dynamic evenness in their playing, which is often said to be a sought-after feature for virtuosic performance.

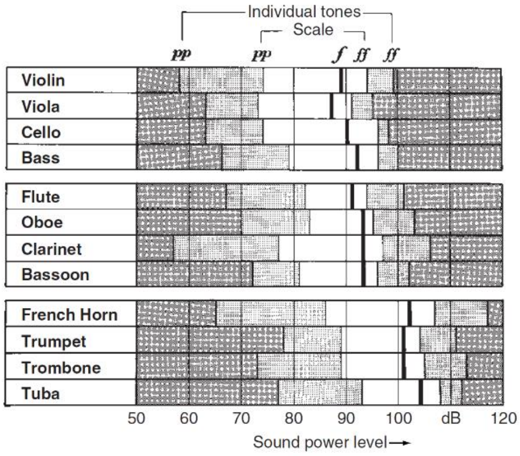
There are also cases in which *pp* written for the trumpet in a high register or *ff* for a flute in a low register is either meant to produce a special effect or else the composer is inexperienced. There are, in my view, still instances where *f* is needed for low flute and *p* for high trumpet, for example, when a *tutti* chord is played and the overall dynamics are meant to be set according to the weakest or strongest instrument. For those cases, the *f* and *p* need to be defined for registers in which it is not possible to produce an actual *f* and *p*. If the dynamic range of the

---

[213] Pedrielli, Carletti, and Casazza 2008, p. 2209.

[214] Krämer 2011.

lowest tone of the flute is just 10 dB SPL, then the difference between *pp* and *ff* in that register is 10 dB. If, for example, there is a musical passage in which the clarinet plays with the flute in the same register, the clarinetist must either set her dynamic range according to the flute or the composer must indicate in the score that the dynamics for each instrument are marked with respect for the instrument's capabilities in the current register. In the Score-Tool project, I see no other way than use the latter option.



Example 26. From Meyer (2009), p. 360. Sound power levels of orchestral instruments corresponding to notated dynamics. The levels are the average of many notes played in the full range of a given instrument. The black stroke indicates the level of *forte* and the light gray stroke indicates the variance in playing the extreme dynamics, *pp* and *ff*. Note how wide the dynamic range is below *forte* compared to the dynamics over *forte* and also how much more variance there is in different interpretations of *pp* than in *ff*.

Aside from Krämer's study (2011), systematic sound power measurements for orchestral instruments across the dynamic register are few. The data can be still collected from different sources, putting together instruments' technical abilities to play dynamics with the measured dynamic ranges of the whole instrument. For example, Meyer introduced a rough graph of dynamic possibilities for orchestral instruments, as shown in Example 26.[215]

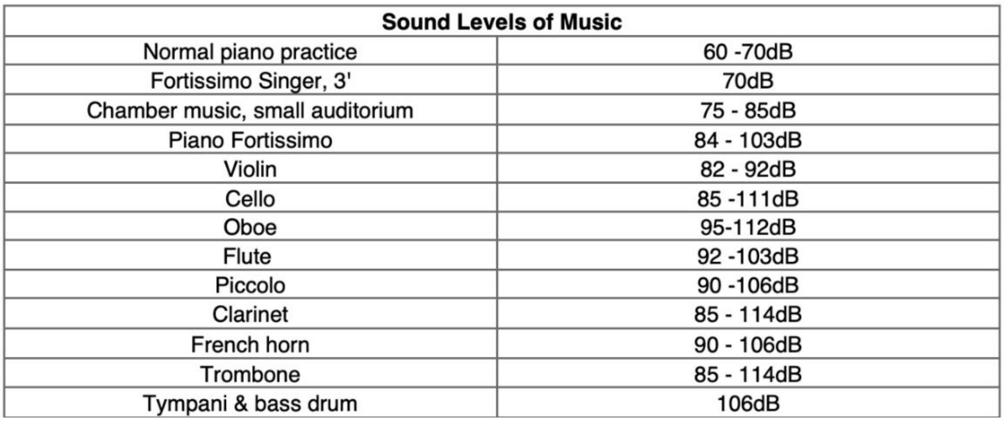
The notable thing in Meyer's graph is that there is a significantly larger dynamic range reserved for soft dynamics than for loud dynamics, especially for the clarinet and horn. This suggests that both clarinet and horn must adjust their dynamic palette according to the rest of the orchestra and perhaps not use their full dynamic potential.

Another interesting source comes from the DPA microphones company, which sells high quality equipment for orchestral recordings. The company has estimated the dynamic range and the typical maximum SPL level for the majority of instruments. For example, the violin's typical max SPL at 3 meters from the player is 95 dB, and the dynamic range is 30 dB.[216] That indicates that a violin's *pp* is around 65 dB, and its *ff* is 95 dB, which is approximately on par with Krämer's measurements.

---

[215] Meyer 2009, p. 360

[216] DPA Microphones 2021.



Sound Levels of Music	
Normal piano practice	60 -70dB
Fortissimo Singer, 3'	70dB
Chamber music, small auditorium	75 - 85dB
Piano Fortissimo	84 - 103dB
Violin	82 - 92dB
Cello	85 -111dB
Oboe	95-112dB
Flute	92 -103dB
Piccolo	90 -106dB
Clarinet	85 - 114dB
French horn	90 - 106dB
Trombone	85 - 114dB
Tympani & bass drum	106dB

Example 27. From the Eastern Kentucky University website 2021, giving approximate decibel levels to which musicians are exposed in rehearsals. The measurements were made with a microphone attached to the musicians’ ears (except for singer, which is measured from three feet distance). The decibel levels are higher than the measurements made in the audience.

Measurements of the health and safety effects on humans of SPL levels of orchestral instruments differ from analytical measurements. The reason may be that these measurements are made, for example, to measure musicians continuous exposure to sound. That is why the SPL levels indicated in Example 27 are measured only a short distance away from the instrument, with a microphone in musicians’ ears. Both the lower and upper limits are greater than in analytical measurements, probably for safety reasons to protect the musicians’ ears. For example, producing 114 dB on a clarinet clearly exceeds the dynamic range of normal orchestral music. A table borrowed from a musician’s health research site is shown in Example 27.[217]

6 MY OWN MEASUREMENTS IN ORCHESTRA REHEARSALS

The orchestra instrument decibel levels in the previous chapter are measured in test conditions that probably do not reflect a real-life situation. Therefore, I did my own measurements in the orchestral rehearsals of my opera, *All the Truths We Cannot See*. I used a calibrated class 1 decibel meter, Sinus Tango, with an accuracy of +-0.7 at 1 kHz. The measured values are A-weighted, i.e., the sensitivity of different frequency areas of the hearing system is taken into account. The measurement position in the concert hall (Sonore Hall in the Music Centre in Helsinki) was on row 8 in the audience, approximately 6-7 meters from the conductor and the nearest player. The measuring distance is beyond the critical distance of the hall. The size of the orchestra was 2222 2210 timp. perc. strings (65432).

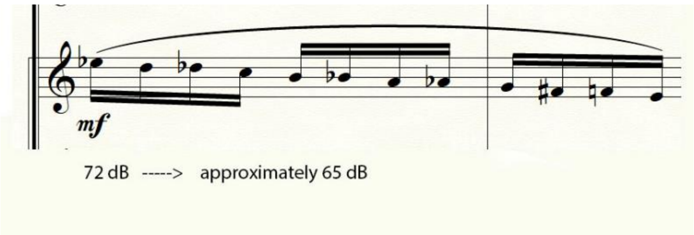
The sound pressure levels at the measuring point, especially in *forte*, were approximately 10-15 dB lower than the dB levels in Example 26. The overall dynamic level was between 55 dB and 92 dB. The softest reading was the *pp* tremolo of the violin section, and the loudest reading was from the trumpet and bass drum playing *f*. The whole dynamic range was thus 37 dB, which is much lower than the range indicated in Meyer’s graph in Example 26. The 30-second average sound pressure level, i.e., the equivalent level LA_{eq}, in passages without high or low

---

[217] Eastern Kentucky University website 2021 (https://music.eku.edu/sites/music.eku.edu/files/ekuhealthandsafety.pdf).

dynamic peaks was between 75-79 dB, which is within the range indicated in Example 26 and Example 27.

The interesting phenomenon in the live performance situation is that, in melodic phrases, the indicated dynamic seems to apply only to the first strong beat of the notated phrase, as shown in Example 28. This seems to be the natural interpretation of the notation system, and if the phrase is interpreted with a truly constant dynamic level, it would sound unnatural. In the rehearsal situation, the decrease in sound pressure was less prominent in the brass instruments.



Example 28. In a test recording in Sonore Hall in the Music Centre in Helsinki with live orchestra. The phrase played by Violin I measured 72 dB on the first strong beat of the phrase and 65 dB in the rest of the phrase, although the dynamics are uniformly marked for the whole phrase.

The readings from the cellos and double bass playing in unison in low register were a surprise. The passage is notated *mf* in the score, but the measured sound pressure level was just 63 dB, which is well below the average *mf tutti* orchestra sound. Part of the low reading is the result of the implemented A-weighting of the dB meter, part of the natural decrease in sound pressure described in Example 28. Part of the reading was the situation, namely, the orchestra was rehearsing an unfamiliar part. Still, the 63 dB sound pressure was surprisingly low, especially given that the bass section was to accompany *mf* violins, which gave readings of about 72 dB.

Another surprisingly low reading was from the left-hand *forte* octaves in the piano. Even though those octaves sound very loud if the piano is in a living room, in a concert hall the reading was just 70 dB. That is much louder than the low reading from *mf* cellos and bass, but still lower than *mf* violins.

The biggest sound pressure difference between registers was found with the French horn, playing a melody *forte*. The readings were about 67 dB in register below C4 and 77 dB in register above C4. This can be explained by both the A-weighting of the SPL meter and the player's increased blowing pressure to produce the high notes on the instrument.

There was also a big difference in the string instrument readings, depending on the playing technique. The SPL level of the violin pizzicato playing *mf* was on par with the *pp* tremolo, only 55 dB. In the cellos, the pizzicato was a bit louder, about 56-57 dB, but still very soft compared to the general orchestra's *mf* reading, which was well above 70 dB. On the other hand, the pizzicato sound has a very sharp attack, which makes it audible even at low sound pressure levels.

In my experience, the harp sound is easily masked by the orchestra, even in the standard orchestral repertoire. In the rehearsals, the harp playing alone in an expressive passage measured 60 dB. This was hardly increased in playing *f*. The 60 dB sound is so soft that almost any concurrent orchestration needs to be marked *pp* if the expressiveness of the harp sound is to be audible.

The highest sound pressure readings came from a single bass drum and timpani strokes. Momentary peaks were about 92-93 dB, which felt really loud. The loudest sustaining notes were played *f* by trombone and trumpet. In both instruments, the reading was between 88 and 90 dB, which was enough to dominate the whole orchestral timbre in *tutti* playing. Also, as stated earlier, the decrease in dynamics was hardly evident in those instruments playing sustaining tones.

7 PREPARING ANALYTICAL DATA FOR THE SCORE-TOOL APP

Today there are dozens of orchestral instrument sound banks available in which each instrument is recorded with good quality. The need for such sound banks comes from the film industry, where sound banks can be utilized to produce music that sounds like a real symphony orchestra and quickly. With a static chord, it is nearly impossible nowadays to tell if the chord is played by a real orchestra or by a computer with an orchestral sound bank. Problems arise when an instrument changes pitch, because the changes would also have to be recorded; otherwise the change sounds too mechanical. In my view, the difference between cheap and expensive sound banks comes from the algorithm used to mimic the acoustical instrument pitch change, not from the recording quality of the instruments themselves.

For the Score-Tool project, the transition data from one pitch to another are not important, because the Score-Tool App does not “play” the score, but instead uses the analytical data on timbre. For the basis of the Score-Tool data, I chose *The University of Iowa Musical Instrument Samples* (MIS), which is a well-documented, completely open source, with quite a comprehensive instrument selection. I added some instruments from another open-source library, *The Versilian Community Sample Library*, which is not so well documented, but has some important orchestral instruments that are missing in MIS, such as harp and timpani. The MIS library contains only raw recordings of instrument sounds; unfortunately, some of the tones are played out of tune and recorded with 3-12 notes played in one file. For extracting individual notes, I created a Python script, which seeks the beginning of each note and removes the silence from the sound files. The total size of instrument samples used in Score-Tool is about 10 Gb. In the Score-Tool analysis, only the analytical data are used, not the samples themselves, which reduce the data amount by approximately 1,000 fold, to 10 MB.

The advantage to using only analytical data in the Score-Tool App is that the analysis of temporal slices of orchestration is not restricted by the sample length.[218] The spectral resolution of the analysis is a trade-off between the temporal and the frequency resolution; therefore, if an analysis is done of a live performance of an orchestral work, the analysis window is restricted by the length of the harmony analyzed in the score. If the harmony is heard only for a fraction of a second, which often happens with harmonies in music, the spectral resolution in the lower frequencies will be poor. In the Score-Tool App, the full resolution spectrum can be used even for harmonies of the shortest duration.

Before the analysis part, each sample is normalized at the SPL level obtained from the table, which is discussed later. The normalization is done in a time domain, so the contribution of all spectral elements to the SPL level is taken into account. An SPL level fine-tuned later in the spectral domain is thus not accurate, because the level adjustments in the spectral domain are hard to translate into the time domain. There are a few missing samples in the normal playing range of the instruments included. Missing samples are transposed from the nearest possible

---

[218] See Part II, section 4.1, “Digital representation of sound waves.”

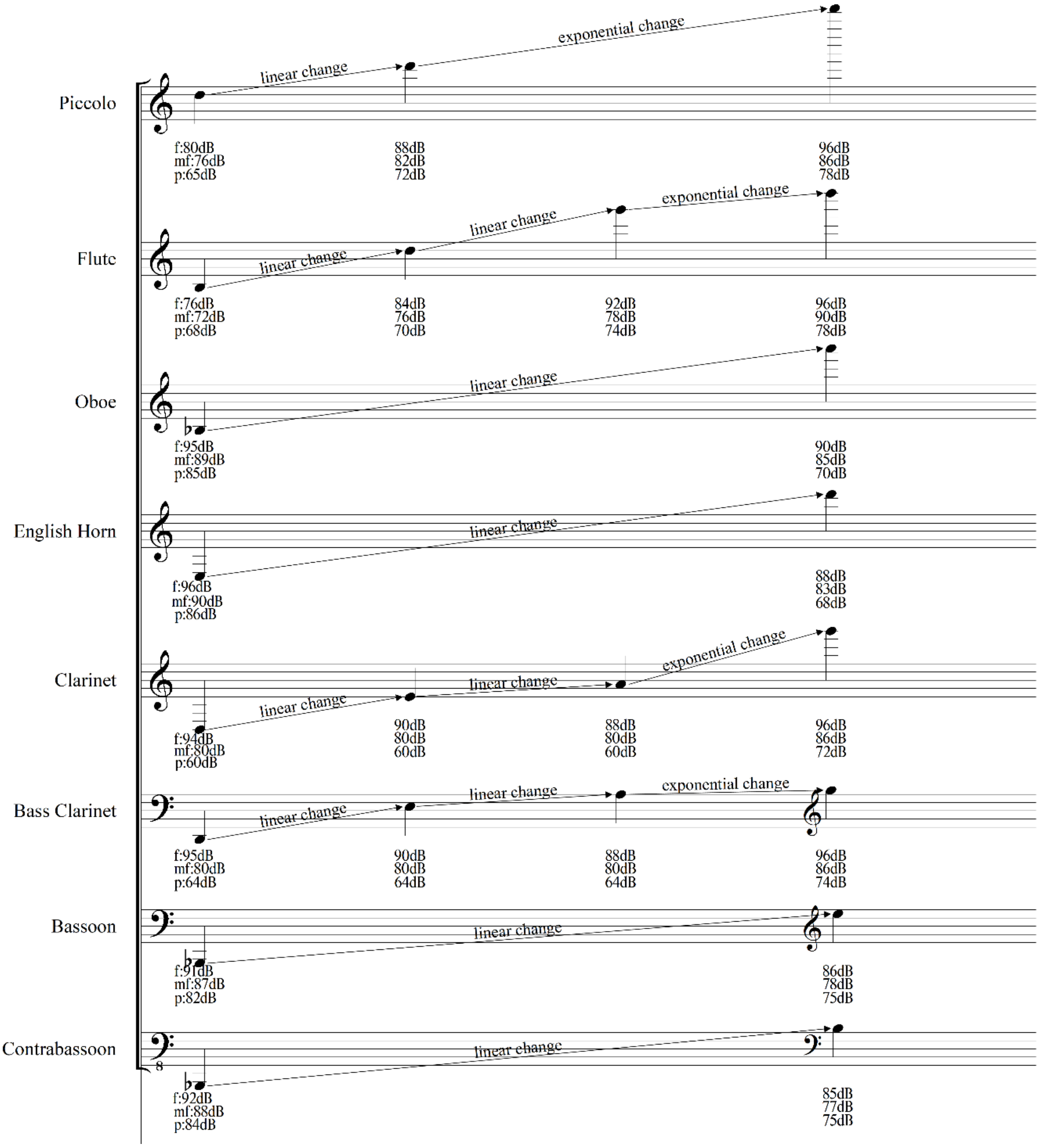
candidate using a resampling method. The largest transposition interval resampled is a whole step.

The Score-Tool database is in the form of a Python dictionary, which is fully loaded into the database memory. Because only the analytical data of each instrument sample are stored, the database is significantly smaller than the audio data. The following parameters of each sample are stored: up to the 30 highest peaks in frequency-amplitude pairs, 12-point MFCC vector, 108-point masking curve, and the spectral centroid.

As stated earlier, only three levels of dynamics are used: *p*, *mf*, and *f*, with a fine-tune possibility. For decibel levels, I created a table that combines data from the orchestration handbooks reviewed in Part I, Chapter 3, and SPL measurements. The table is tailored specifically for each instrument and can be adjusted later, but then the whole database would need to be recalculated.

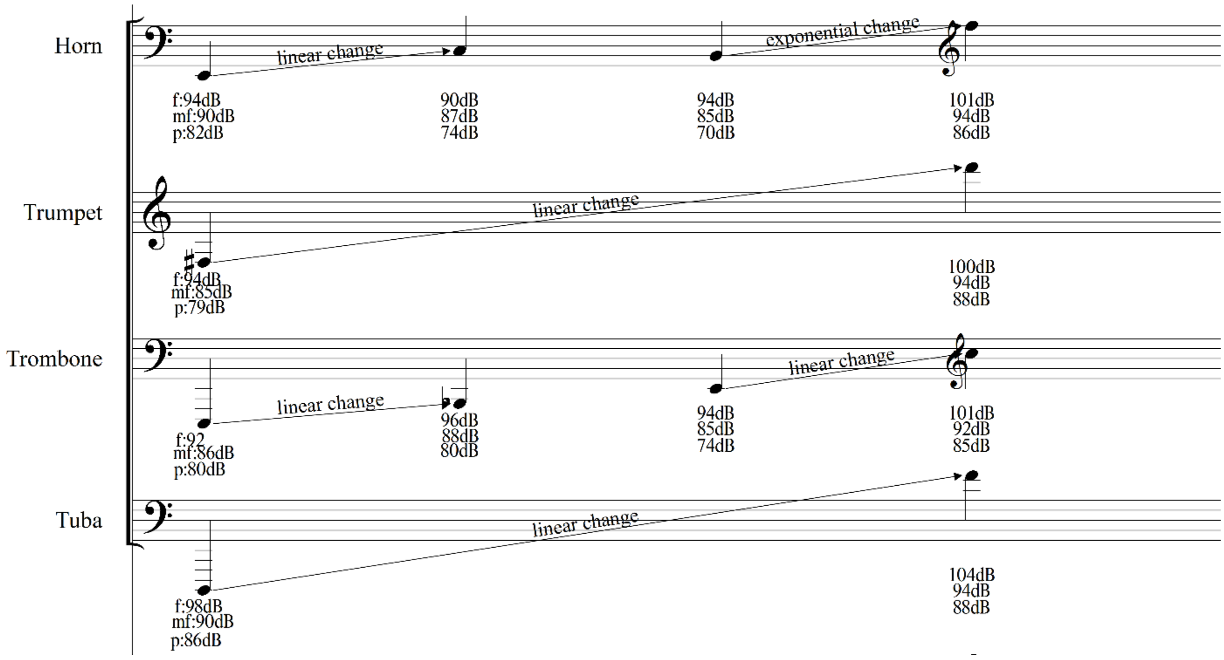
In the following examples, there are SPL levels of *p*, *mf*, and *f* marked for orchestral instruments. The levels are marked at both ends of the pitch range and to the pivot points of the instrument's register, if there is one. The line connecting the points indicates whether the transition of the SPL levels is "linear" or "exponential." In transitions, the dB levels increase or decrease according to the function of the chromatic pitches. In "exponential" transition, the transition speed is first slow, but increases towards the goal value, such as in a trumpet *crescendo* from *p* to *f*. In "linear" transition, the transition speed is constant, such as in a clarinet *crescendo* from *p* to *f*.

Example 29 shows my interpretation of the SPL levels of woodwind instruments at sounding pitch, with the technical possibilities of each instrument taken into account.



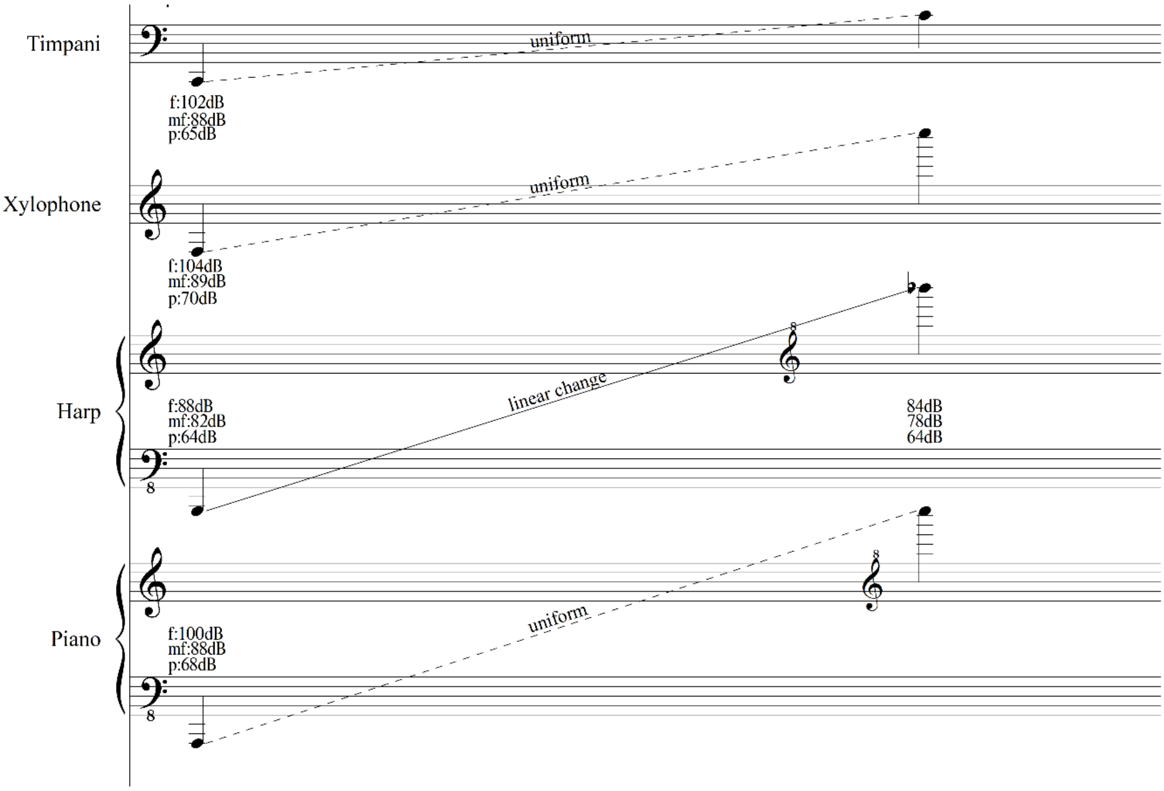
Example 29. Sound pressure levels of woodwind instruments in the orchestra

Example 30 shows the SPL levels for brass instruments.



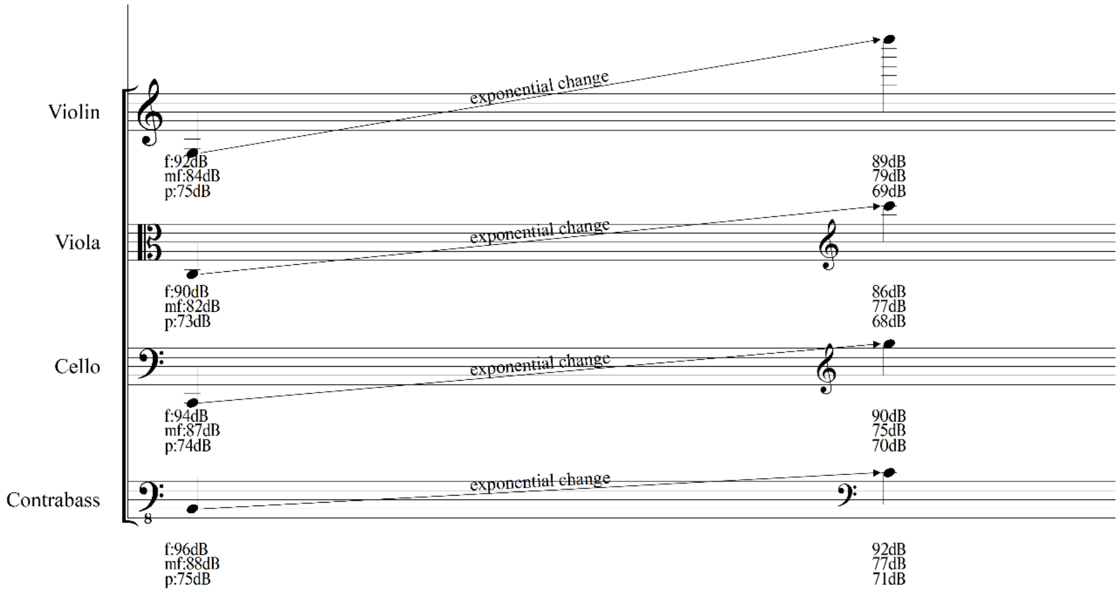
Example 30. Sound pressure levels of brass instruments in the orchestra.

The trombones' lowest range consists of pedal tones. Most of the trombones have an extra valve for "missing" tones between the pedal and the first harmonic, which then have the dynamic range of the first harmonic tones. In Example 31 are SPL levels for percussion, piano, and harp.



Example 31. Sound pressure levels of timpani, xylophone, harp and piano in the orchestra.

The levels of the piano are uniformly, although the highest strings actually have less power than the lowest because of their shorter length. The difference has not been documented in any published paper, so the uniform level range is used. Lastly, SPL levels for strings are shown in Example 32.



Example 32. Sound pressure levels of string instruments in the orchestra.

The upper range of each string instrument may vary according to a player's skills. The solo and section levels are not separated because the extreme dynamics are not used. The database also includes solo strings, which, despite matching the SPL level in the string section, have individual samples, and thus a non-matching spectral content.

These are the SPL levels used in the Score-Tool App. Currently, the code that assigns the SPL levels and does the analysis is not part of the Score-Tool App. In the future, my idea is to make it possible for users to add instruments to the database, but the SPL levels of corresponding dynamics must be carefully set, with the reference to existing measurements. Otherwise, the usability and reliability of the App may suffer.

The most usable measurement data for my purpose is a combination of analytical and technical measurements, which give an estimate of the SPL level several meters away from a player in the concert hall. This does not take directionality into account, because the critical distance, especially calculated with a directionality coefficient, would most certainly be greater unless the hall where the measurements were taken was very small. Nevertheless, using this data gives a tolerable margin of error for the Score-Tool App, and the data can be updated if new and more accurate measurements are available in the future.

8 ANALYSIS MODELS ADAPTED AND DEVELOPED FOR SCORE-TOOL

8.1 MPEG LOSSY CODING

MPEG audio coding includes a model based on psychoacoustics, namely, a computational formula describing the perception of sound in the outer, middle, and inner ear.[219] The phenomenon behind the model has been known since the early decades of the twentieth century, but the subject became relevant only in the early internet era when audio coding was introduced. The first popular digital audio formats, such as DATs and CDs, used 44.1– 48 kHz sampling rate and 16-bit samples per channel, which in the case of CDs corresponds to 1.4 Mbit of data per second. In the early internet era, the connections were slow and digital storage expensive, so efforts were made to find ways to compress the digital audio without losing the high-quality sound.

Digital music audio contains many redundancies, which can be exploited to compress the file to half its original size without losing any information. This is called lossy coding. The human auditory system, however, has many limitations. The lossy coder exploits those properties and often can make files 10 to 12 times smaller without perceptual loss.[220]

The basic mechanism of a lossy coder is to reduce the bit rate of the digital audio. The bit rate reduction causes quantization noise, which without the use of a model based on psychoacoustics, disturbs the listening experience and causes the audio to lose its “high quality rating.” To keep the perception of the audio quality high, the audio stream is divided into small chunks of data, which are converted into the spectral domain, furthermore the spectrum is divided into frequency sub-bands and the amount of masking created by the loudness of the music on each sub-band is calculated. Next, the spreading of the masking to adjacent bands is determined, and finally the signal-to-mask ratio in each sub-band is obtained for the given chunk of data. While the amount of quantization noise created by reducing the bit rate is known, the noise can be shaped using the obtained sub-band signal-to-mask ratio, so that the noise is not audible to the listener. For example, if the music has a quiet and sensitive section, perhaps only 1-2 bits in a few bands are reduced, and in the climatic section with loud percussion and brass, the codec uses perhaps only 1 bit in nearly every sub-band. Reducing the bit rate to 1 reduces the file size by a factor of 16. The most well-known product of using this model is mp3 compressed music, popular at the end of the 1990s, yet it gradually lost popularity as better coding methods became available and greater data transfer speed and storage capacity were developed. The mp3 refers to the coding standard MPEG-I Audio layer 3, which defines the method for coding and decoding using the psychoacoustic model.

The MPEG psychoacoustic model has been widely tested, one could say by billions of people, since the discussion about whether the quantization noise in mp3 files is audible or not has taken place in multiple forums for nearly two decades. In general, it could be said that the MPEG coding does not disturb the perception of music for the everyday listener, but with concentration and practice, the disturbing noise can be detected. The MPEG coding model is thus a usable and adequate model, proven to correlate well with the listening experience.[221]

---

[219] Noll 1997, p. 62.

[220] Lincoln 1998.

[221] Kirby and Watanabe 1996, p. 35.

8.2 WHY I WANT TO USE THE MPEG PSYCHOACOUSTIC MODEL

I decided to use the MPEG model in my project both because it had been widely tested and for its computational efficiency. However, there are a few issues in using this model to determine the audibility of an instrument in orchestral texture. The biggest is that the original model was developed for tone-masking-noise cases, and in my project most of the cases are of tone-masking-tone type. Noise is easier to mask than a tone because noise does not interfere with tonal components. In tone-masking-tone cases, possible roughness and beating caused by close tonal components reveal the presence of a masked tone, even though the tone itself might be inaudible. This is, however, a limitation I am ready to accept, because in determining the audibility of an instrument, the question is not whether we barely notice the interference caused and say “aha,” but that we can actually hear the music the instrument plays. I thus modified the MPEG model to suit my needs, which is described below in section 8.3.

8.3 HOW THE MPEG PSYCHOACOUSTIC MODEL WORKS

The steps for calculating the MPEG model, according to the ISO/IEC standard are as follows:[222]

Step 1 - Calculation of the FFT for time to frequency conversion.

Step 2 - Determination of the sound pressure level in each sub-band.

Step 3 - Determination of the threshold in quiet (the absolute threshold, the hearing level).

Step 4 - Finding the tonal (more sinusoid-like) and non-tonal (more noise-like) components of the audio signal.

Step 5 - Decimation of the maskers to obtain only the relevant maskers, meaning, practically speaking, sinusoid-like components.

Step 6 - Calculation of the individual masking thresholds.

Step 7 - Determination of the global masking threshold.

Step 8 - Determination of the minimum masking threshold in each sub-band.

Step 9 - Calculation of the signal-to-mask ratio in each sub-band.

8.4 THE PSYCHOACOUSTIC MODEL IN THE SCORE-TOOL APP

Step 1 – The Fourier transform is calculated with 512 or 1,024 samples. This is because music has rapid changes in time; if more samples are used, the spectrum obtained would contain components that are not heard simultaneously, but in sequence. Using more samples has the benefit of increasing the spectrum resolution in lower frequencies. In my project, I use 44,100 samples in a Fourier transform; using real instrument sounds without rendering the actual music avoids unwanted tones in the calculation. The spectrum is always exactly the one in the score. The Fourier transform could in theory be as long as the instrument sound sample, but for further processing it is convenient to have the same length for all transforms. In the first step, the audio is normalized to the reference level of 96 dB SPL, which I omit because the normalization would destroy the carefully set dynamic levels of instrument sounds. Before the Fourier

---

[222] ISO/IEC 1996, p. 85.

transform, the Hanning window is applied to smooth the resulting spectrum. The Hanning window in the ISO/IEC standard, where N refers to data points in the sample, is as follows:[223]

$$h(i) = 0.5 * \left\{1 - \cos\left[\frac{2\pi(i)}{N}\right]\right\} \quad 0 \leq i \leq N - 1$$

Step 2 - Step two is applied as in the original model.

Step 3 - The threshold in quiet is determined, the curve obtained from the table in the ISO/IEC standard appendix. The source of the curve is not defined in the ISO/IEC Directive, but the source resembles the equal loudness curve for hearing threshold in quiet.

Step 4 - The spectral components are categorized according to tonal and non-tonal components. The idea is that non-tonal components are already noise-like and thus do not contribute to masking curve calculations. This is important only in tone-masking-noise cases. Calculation of instrument noise components, such as attack transients and bow noises, is omitted in this version of Score-Tool to save computation time.

Step 5 - In the original model, components closer to each other than half the critical band are decimated. In the standard model, this is done to reduce the number of maskers considered for calculating the global masking threshold.[224] The decimation procedure keeps the component with the highest power and removes the smaller components from the list. For my project, this decimation procedure is omitted, because although it does not say so in the ISO/IEC Directive, the procedure is likely carried out to avoid summing up components in the same critical band. In my project, I omit the decimation and sum up each component to obtain the "real" excitation value for each critical band.

Step 6 - Applying the masking and spreading model to the spectrum. The masking and the spreading of the masking is calculated according to the spreading function, which is defined in the mathematical section of this report (Part II, Chapter 4). In this step, the tonal and non-tonal maskers are treated differently, but in Score-Tool, I use only formulas for tonal maskers, even though some of instrument spectra, especially the percussion, certainly contain many non-tonal maskers. Non-tonal maskers would have less masking effect than tonal maskers, but differentiating the two would require additional calculations, which would slow down the performance of Score-Tool. My decision to treat non-tonal maskers as tonal results in some masking curves showing a bit higher masking in some cases than the original MPEG model, which I do not find problematic.

In Step 6, the masking curve is obtained, i.e., the curve in the frequency domain describing the amount of masking that the current time frame creates in our hearing system. In the ISO/EIC standard, the output is a value for 108 individual frequency points, with a table indicating the corresponding fraction of a Bark band for each point.

Steps 7, 8, and 9 – The bands are combined into 32 uniformly distributed areas in frequency space, and the minimum masking threshold and the calculation of the signal-to-mask-ratio are calculated for each of these 32 bands. In my case, it is more convenient to keep the 108 bands with the information on Bark bands obtained in Step 6, because the 32 uniform bands are too

---

[223] ISO/IEC 1996, p. 86.

[224] ISO/IEC 1996, p. 88.

“rough” a resolution for orchestral music. For audio coding, this “roughness” is included to play it safe (32 uniform bands cover very well the 16 non-uniform Bark bands used for the ISO/IEC standard paper’s 24 kHz music resolution), but for my needs the greater resolution is useful, and the conversion to Bark bands is done in the final step in determining the audibility.

The table of critical band rates and the absolute threshold in the ISO/IEC standard is created for a sampling rate of 24 kHz, resulting in the highest frequency in question being 12 kHz. As stated also by Terhardt (1981, p. 680), the aurally relevant frequency region is about 20 Hz to 5 kHz, so the table covers enough frequencies for the masking calculations for music.

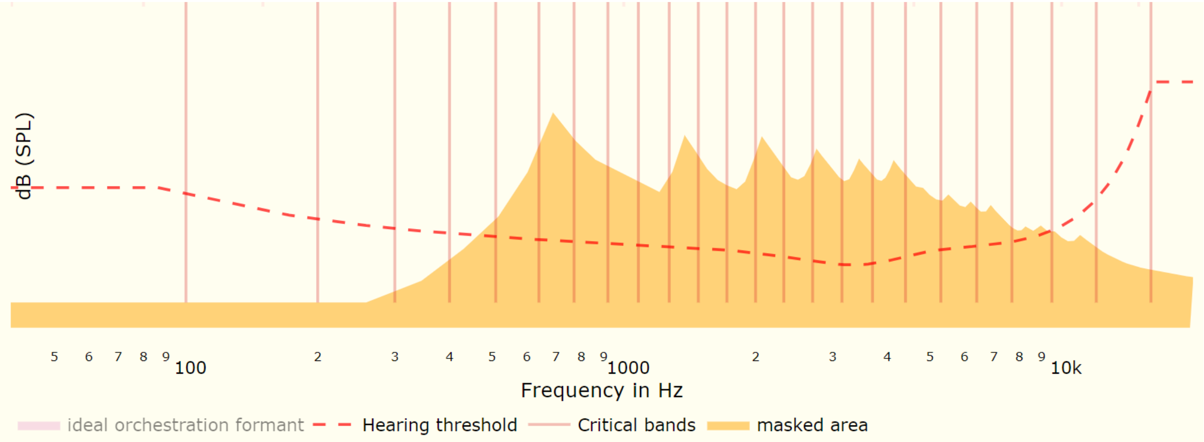
The ISO/IEC standard presented here is implemented and used in the Score-Tool App with the slight modifications stated above.

8.5 TARGET AUDIBILITY PREDICTION IN SCORE-TOOL APP

Composers using the Score-Tool App in the first testing phase found it difficult to interpret all the data in the graphs and asked if there could be a simple indicator to show whether the target was audible or not. Therefore, I introduced the *audibility prediction* algorithm. This algorithm takes the masking, blending, and timbre similarity measurements into account along with the weighting set based on my experience gained during this project and outputs a single value indicating the likeliness of a target’s audibility.

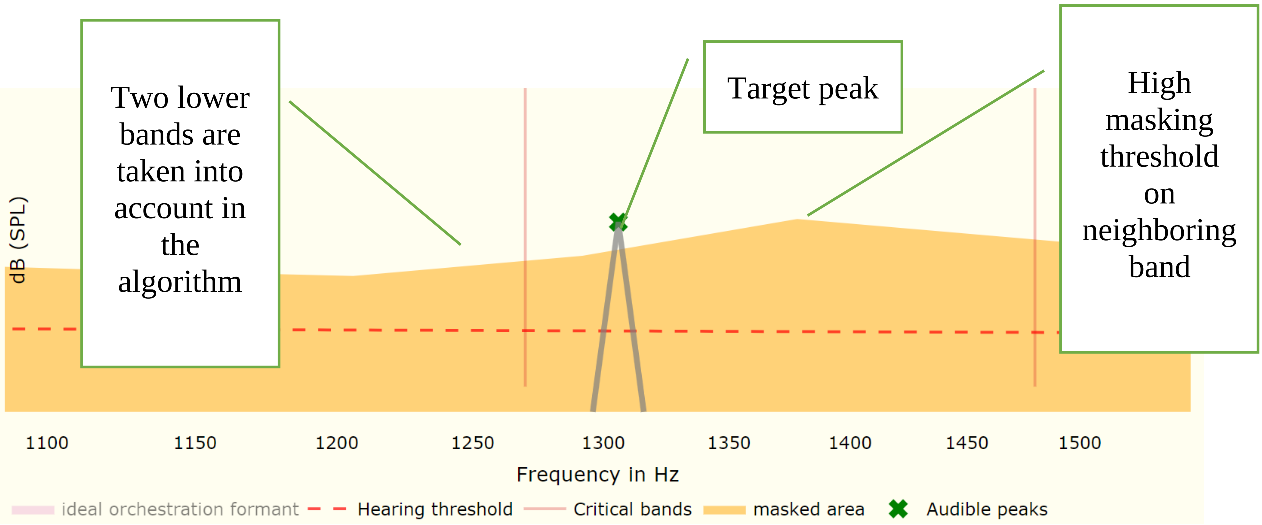
The masking curve alone does not give enough information about the target’s audibility. This is because the masking curve is calculated using the MPEG coding formula, which was developed to determine whether the quantization noise, i.e., an inharmonic buzzy sound, is audible or not. In orchestration, at least for me, it is not enough that the target is just barely audible. Composers might want the target to be clearly distinguishable.

The masking curve consists, as described above, of a masking threshold value in dB on 108 frequency bands. Each band represents roughly one-fourth of a Bark, i.e., a critical band, as shown in Example 33. Thus, a sharp peak on one of the 108 frequency bands results in masking neighboring bands because that is the width of the critical band in the human hearing system. Showing the masking threshold only on Bark bands is not the right solution, because the Bark frequencies are only estimations, and target peaks very near the border of a Bark frequency would yield false results in the calculations.



Example 33. In the Score-Tool App, the masking curve is presented as a masking threshold value in dB on 108 frequency bands. Each band represents roughly one-fourth of a Bark, i.e., a critical band, shown with vertical lines in the figure. This is a masking curve produced by a

violin tone F5. The "pyramids" come from the spreading function applied to each of violin partials.



Example 34. The Score-Tool masking algorithm takes into account the maximum of two lower, one upper, and the current masking threshold values on the target frequency band. In the figure, one upper band shows a higher masking threshold than the target peak, which results in the target peak being under the masking threshold in the audibility calculation.

In Score-Tool, when a target peak is compared to the masking threshold, my algorithm takes the maximum of the two lower, one upper, and the current masking threshold values on the target frequency band, as shown in Example 34. This eliminates the border-case problem with Bark bands, because the peak would always be in the "center" of the assumed critical band.

My intention is for the results to reflect the sounding experience of a live performance. Thus, the masking curve cannot be treated as an absolute value; an instrumentalist can play a few decibels louder or softer or alter the playing technique slightly in the performance, which affects the masking threshold. Therefore, I decided to introduce different thresholds that affect the audibility percentage as seen in Example 35.

	All peaks	Over 2 peaks	1 peak
Peaks over 15 dB above masking threshold	100% audibility	95% audibility + percentage of peaks above the threshold	80% audibility + percentage of peaks above the threshold
Peaks over 10 dB above masking threshold	100% audibility	80% audibility + percentage of peaks above the threshold	50% audibility + percentage of peaks above the threshold
Peaks over 6 dB above masking threshold	90% audibility	50% audibility + percentage of peaks above the threshold	30% audibility + percentage of peaks above the threshold
Peaks over 0 dB above masking threshold	40% audibility	30% audibility + percentage of peaks above the threshold	20% audibility + percentage of peaks above the threshold

Example 35. The table shows the initial calculation of the percentage of masking in the Score-Tool App. In addition to these values, the spectral centroid and timbre distance to orchestration affect the audibility percentage.

In addition to this masking percentage value, I realized, especially in testing the Score-Tool algorithm with my horn concerto *Sonority* (described in Part III, section 3.1), that a low spectral centroid of the target and a close timbral distance of the target to the orchestration also affect audibility. Thus, I implemented a gradual decrease in audibility percentage when the spectral centroid falls under 2 kHz, which I have set as a threshold of “dark” timbre and which blends easily into the orchestration (described in Part I, Chapter 7). The formula I decided to use is the following:

Initial audibility percentage - $\left. \frac{2000}{target\ centroid} \middle/ 5 \right. * 100$

With this formula, the target with a 2 kHz spectral centroid value results in a 20% decrease in audibility. If the target centroid were to be 1420 Hz, the decrease would be 28% percent. The formula is the result of my experience in comparing the audibility values of the Score-Tool with live orchestral performances. I continue to make adjustments to the formula as I acquire more experiences with orchestration.

In a live performance, a target’s close timbral distance to the orchestration did not decrease the audibility as dramatically as the target’s low centroid did. Therefore, I decided to use a simple formula with only one threshold value: If the target’s timbral distance from the orchestration is under 40 on Score-Tool’s scale, then audibility is decreased by 10 units. This is again a rough estimate based on my experience and can be changed according to further experiments.

In the Score-Tool App, the audibility percentage is color-coded in the score. An audibility percentage of 100 is shown in a green color, which fades gradually to yellow and closing at

75%. A value of 75 results in a yellow color that fades to red closing at 50%. A value of 50% results in the red color fading to brownish red closing at 0 %. This scale is shown in the Score-Tool App in Example 36.

Color indications: Target masked Target nearly masked Target audible

Example 36. Colors indicating the audibility of the target. Green indicates the value 100, i.e., the target is completely audible. Yellow gradually turning to red indicates a value of 50, and red indicates the value is near zero, i.e., the target is in danger to be masked.

As the colors suggest, green means that audibility is good, yellow means that some masking might occur, and red means that there might be serious audibility issues.

In conclusion, the audibility percentage is not the "amount of sound audible," as the percentage unit might suggest. It is a rough estimate of a person's chance to hear the target sound on a scale that is partly based on my subjective experience. Because of this, in a live performance, the Score-Tool audibility value of 55% might or might not be better than a value of 45%, but a value of 80% is definitely better than a value of 20%.

8.6 HOMOGENEITY OF ORCHESTRATION

In this section, I discuss the orchestrator's need and potential uses for a homogeneity measure. I also introduce the means to give homogeneity a numerical value with a method borrowed from data science. I point out the benefits and caveats for using the data science method for timbral data and explain how this method is implemented in the Score-Tool App.

8.6.1 WHY DO WE NEED TO MEASURE THE HOMOGENEITY?

For a composer or orchestrator, the need to know the homogeneity of the orchestration can be related to technical or artistic reasons. From a technical point of view, there might be a desire to obtain the maximum possible blend within orchestral layers. And because similar timbres tend to blend better than contrasting timbres,[225] homogeneous orchestration is an advantage. Another technical need might arise in determining the audibility of the target with the Score-Tool method. The homogeneity parameter can be used to predict whether the target has a timbre that is distinct from the orchestration. If the orchestration is highly heterogeneous, then the target timbre, despite the high audibility masking-wise, could be hard to detect among the high variety of timbres, i.e., in the heterogeneous orchestration.

8.6.2 ARTISTIC NEEDS

The need for a homogeneity parameter can also be artistic. For example, in orchestrating a passage, I have often felt the need to compose a chord with a highly homogeneous sound. Homogeneous sounds may be needed, for example, to shift the focus of the music from orchestration to harmony. The need can also be the opposite: to attract a listener's attention with heterogeneous orchestration when there is not much happening in the harmony.

The concept of timbre homogeneity appears every now and then in casual discussions with colleagues, but not necessarily with this exact term. I often hear arguments of this kind: "I like composing for string quartet because of its uniform sound color"; "It's easy to compose for monochromatic ensemble"; "Composing for wind band is hard because of the heterogeneity of

---

[225] Lembke and McAdams 2015.

the timbre.” I agree with these views. And the problems they bring up become even more complex with a full orchestra.

8.6.3 TIMBRE HOMOGENEITY IN ORCHESTRATION HANDBOOKS

The homogeneity of timbre is rarely addressed as a parameter in orchestration handbooks, orchestration literature, or even in teaching orchestration. The timbre variety within an ensemble could be thought of as part of a composer community’s tacit knowledge. The ability to orchestrate homogeneously or heterogeneously is acquired little by little through experience and in discussions with more experienced colleagues.

8.7 COEFFICIENT OF VARIATION

In the Score-Tool App, I introduce an algorithm borrowed from data science that makes it possible to check the homogeneity of the orchestration. In data science, especially in business-related data science, researchers try to find formulas that would predict future developments. One sub-class of this practice is the demographic approach: a study of the relationship between kinship structures. The most frequently used measure of demographic heterogeneity is the *coefficient of variation*.[226] The mathematical definition of the coefficient of variation, or C_v can be found in Part II, section 4.9.

The reason I became interested in using the C_v method in Score-Tool is that the C_v measures the variability of a series of numbers independent of the unit of measure used for the numbers.[227] In the Score-Tool App, I use the MFCC vector as a measurement of timbre. The MFCC values are unitless. Thus, using the C_v formula to address MFCC would be a natural choice.

8.8 COMPARISON OF MFCC VECTORS

In the Score-Tool App, the MFCC timbre data are obtained from both the whole orchestration timbre and from the timbres of individual instruments participating in the orchestration. This makes it possible to explore the orchestration’s timbral components in relation to the overall timbre. In the MFCC of the overall timbre, there is no way to determine its homogeneity. The only option is a comparison of the timbre components, i.e., the MFCC components. This part of the App is experimental and needs further testing to get reliable results.

Comparing an orchestration’s MFCC components can be done in several ways since the data are consistent. Every MFCC vector consists of 12 values (the first value is omitted), and there are no missing data, because the MFCC algorithm always outputs all values, even if the correlation is zero.[228] The usual methods for comparing consistent data are *mean, variation,* and *deviance*. In MFCC values, data have no units and likewise no actual scale, because the result comes from correlation. For example, a characteristic of the variance (or equivalently, the standard deviation) is that variance is sensitive to the scale on which the variables are measured: if all values are multiplied by a constant c, the variance will increase by a factor of

---

[226] Sørensen 2002.

[227] Abdi 2010.

[228] See Part II, section 4.6.

c as well. One solution to this problem is to use the coefficient of variation (C_v).[229] In machine learning applications, C_v indicates the constancy aspects of the system (data).[230]

The mathematical definition of C_v is simple: it is the standard deviation divided by the mean. The simplicity is also an advantage for me because the computational part of the Score-Tool App is in danger of becoming too heavy. Luckily, C_v supports designing computationally efficient single-pass algorithms, thanks to its elegant algebraic properties.[231] With MFCC, however, there is a minor concern about the shape of the data, because the values of the measurement used to compute the C_v are always assumed to be positive or nil.[232] Furthermore, the coefficient of variation is not definable when the mean is zero, and it will be unbounded when the mean approaches zero. [233] In other words, there is a danger of poor values if the mean is low.

8.8.1 SETTING THE SCALE FOR MFCC VALUES

The MFCC values are negative or positive, depending on whether the *cepstrum* correlates to the cosine wave. The negative values are not valid to be used in the C_v algorithm. There are similar cases in data science, where, for example, only data with positive values are used, and other methods are applied to data not suitable for the formula. One possibility is to convert the data to a scale with only positive numbers. Bindu et al. showed in their research that, while the scale of the data has no effect on C_v, conversion of the data influences C_v exponentially.[234] This can be avoided if the data are normalized, because translation and scale have no effect on normalized data. Bindu et al. also advise avoiding the non-existence of C_v; they recommend coding in a strictly positive zone. Further, they recommend bringing the range of normalization to [1, 2].[235] In my approach, I translate the MFCC values to a positive zone and apply normalization. For example, if I use this method for 5 value vector [-2, 2, 0, 1, 0], normalization to [1, 2] means that the lowest value -2 becomes 1 and the highest value 3 becomes 2, and the rest in between in proportion like this: [1, 2, 1.5, 1.75, 1.5]. This the data which is always valid for C_v formula.

8.8.2 RESULTS AS PERCENTAGES

The values obtained with the formula are usually fractions. In the research literature, C_v is often presented as a percentage, which is obtained by multiplying the value by 100. I also use the percentage presentation because it makes the value more readable in comparison to the fractions. The main thing to remember is that the C_v can also be over 1, resulting in a "percentage" over 100, which may confuse the Score-Tool user. In the App's code, the value is therefore restricted, making every value over it result in a maximum of 100% since values over 100 are automatically "highly heterogeneous."

8.8.3 SETTING THE SCALE

---

[229] Sørensen 2002.

[230] Bindu et al. 2021.

[231] Bindu et al. 2021.

[232] Abdi 2010.

[233] Bindu et al. 2021.

[234] Bindu et al. 2021.

[235] Bindu et al. 2021.

The final step in determining the homogeneity of orchestration by comparing MFCC vectors is setting limits for homogeneity or heterogeneity. For a reference, Bindu et al. (2021) provide in their study a set of values that fall into a certain category of data consistency. In the case of MFCCs, consistency is interpreted as homogeneity. For homogeneity, the scale must be inverted, because the lower value means more homogeneity. Below, I have put the scale provided by Bindu et al. (2021) on my inverted values. On the right-hand side of the table, I have added the attributes I use in the Score-Tool App to determine timbre homogeneity:

Coefficient of variation	Attribute	Attribute translated to Score-Tool
0-5	highly consistent	highly homogeneous
5-15	moderately consistent	moderately homogeneous
15-33	weakly consistent	weakly homogeneous
33-66	weakly inconsistent	weakly heterogeneous
66-100	moderately inconsistent	moderately heterogeneous
>100	highly inconsistent	highly heterogeneous

In Score-Tool, the homogeneity is explored against the user's experience in orchestration, and is to be used for artistic inspiration in experimenting with orchestration.

8.9 CONCLUSION ABOUT TIMBRE HOMOGENEITY

As a final word about the subject, there are alternative ways to measure homogeneity. I chose the C_v option because, for me, it gives expected results in many cases. There has been criticism of using the coefficient of variation, with some stating that it may lead to incorrect conclusions about empirical phenomena.[236] Therefore, I strongly suggest using this tool with enough orchestration experience to interpret the homogeneity value correctly.

9 DATA VISUALIZATION

In this chapter, I discuss the problems with viewing orchestration data on graphs and offer some solutions. I point out the similarities between statistical and musical concepts and discuss the possibilities for borrowing features from statistical graphs for the orchestration-oriented Score-Tool App. I also use examples of graphs in the Score-Tool App to discuss the visualization choices I have made in presenting data to musicians.

The reason I want to borrow visualizations and algorithms from statistics and data science is that the orchestral score itself is not a sufficiently good graph to visualize the central concepts

---

[236] Sørensen 2002.

in the Score-Tool project, namely, masking and blending. In fact, the score provides very little information about those concepts unless the score is interpreted by an experienced orchestration specialist.

9.1 FUZZY DATA

The Score-Tool App, where mathematical algorithms are applied to orchestration, results in a huge amount of numerical data. The data are partly interpreted by the Score-Tool App and given to the user as text output, such as "Orchestration is brighter than target" or "Consider lowering the dynamics of the trumpet." The results also include fuzzy data whose interpretation depends on the artistic needs of users and cannot therefore be set into algorithms. The interpretation of the fuzzy data is therefore left to the artist, but the numerical data must be presented in a form that the artist can read.

In a way, the whole process of orchestration is a task in which artistic choices are based on uncertainty and fuzzy data. There is no one best option for orchestrating a passage, and different choices cannot be placed on a scale where one choice, for example, is 10% better than another. In the compositional phase, I used to make orchestration choices intuitively, based on what I then knew about the orchestra. With experience gained from performances and rehearsals of my works, I became more secure in my choices, but also restricted, because I subconsciously chose to use the kinds of orchestration that had worked for me before. This kind of orchestration resembles the way a child learns about the world: If broccoli eaten with ice cream tastes bad, then do not eat anything green and tree-shaped with ice cream ever again, although something green might taste good with chicken. Similarly, a bad orchestration choice at the rehearsal of one piece might be a good choice for another piece.

One of the main concepts of statistics is data visualization, because anomalies and patterns are easier to recognize from a picture than from a number matrix. Through visualization, abstract or complex information can become obvious.[237]

9.2 STATISTICAL DATA VISUALIZATION

Next, I discuss the basics of visualizing statistical data and examine a musical score as a statistical graph. I also explain the benefits of showing orchestration data visually.

Orchestration may not be directly comparable to statistics, since a statistically orchestrated composition, where, for example, every note is assigned to the loudest possible instrument that can produce the pitch, would probably not sound good and would result in a dysfunctional score. Statistical methods are used in the Score-Tool App, although not to orchestrate, but rather to analyze the existing orchestration. Some basic statistical tools are used in the App, such as mean, deviation, and distance algorithms. So are data visualization techniques, which provide information from the orchestration that has not been previously available because these techniques were invented in this Score-Tool project. In statistics, new kinds of visualizations are constantly being created, because new kinds of data require new kinds of analytical tools.[238]

The visualization must be chosen according to the type of data to be visualized. For example, in order to compare two sets of data, it helps considerably if the scales and ranges of units are the same for both sets in the graph. There are at least four categories to consider in deciding on

---

[237] Wegman and Solka 2004, p. 539.

[238] Wegman and Solka 2004, p. 30.

the best type of visualization: distribution, relationship, composition (or amounts), and comparison.[239] One category does not exclude another, and a graph can describe, for example, distribution and comparison at the same time.

A musical score is also a visualization of musical data. From a statistical point of view, the score describes both the composition of the musical elements and the distribution of notes over time. A musical score is a very good visualization tool for those purposes, since, based on the visualization, the music can be reproduced simply by interpreting the graph (the score). The score is not as good for visualizing relationships and comparisons, which can be seen in the fact that music theorists disagree about those aspects in particular. For visualizing musical relationships, for example, a different kind of graph is needed, but a different graph does not necessarily mean a new design, only a new perspective. Two graphs with the same units can be used to show different aspects of the data. A good example of this is a Schenkerian graph, which resembles the musical score to which it pertains, but is concentrated on showing the harmonic relationships over measures rather than note distribution over time. Therefore, in the Score-Tool App, I thought carefully about the visual design as well as the visual perspective of the graphs.

A musical score serves at least two purposes for the musician: to present and explore musical data. These are also two main reasons for using graphs in the first place.[240] Presenting data is a relatively objective task, but exploration, that is, using graphics to find information and generate ideas, is a much more individual matter.[241] The ideas that a score generates for a musician can be called an interpretation, and the ideas a score generates for a music theorist constitute an analysis. To repeat the point made at the beginning of this chapter, data conveyed by a score are fuzzy. There is not just one way to interpret a score, but unlimited ways.

In a similar fashion, the Score-Tool App's orchestration data are fuzzy and thus open to different interpretations. The primary purpose of the Score-Tool graphs is to suggest orchestration ideas rather than to present an exact perceptive model of the orchestration.

9.3 AUDIBILITY VISUALIZATION MODEL

My approach to the visualization of the data in the Score-Tool App mimics the purpose of a musical score. The graphs are not there to give the final word on masking and blending, but are meant to present these aspects of orchestration and help the user explore them. The graphs are intended for music professionals for the purpose of providing structured and categorized information about the orchestration. Most of the graphs are weighted towards the distributional presentation, showing how the spectral content of the orchestration sound is distributed in our hearing system. The main distribution graph, the masking curve, is weighted with the algorithm that filters the data through the mathematical model of the human hearing system.

Here, I present some examples of the graphs I designed for the Score-Tool App and discuss visualization and the purpose behind the choices I made. Concepts are borrowed from both statistical graphs and from musical graphs. Combining these two should give the educated musician a good view of the masking and blending properties in orchestration.

---

[239] Wilke 2019, p. 37; Alam 2020.

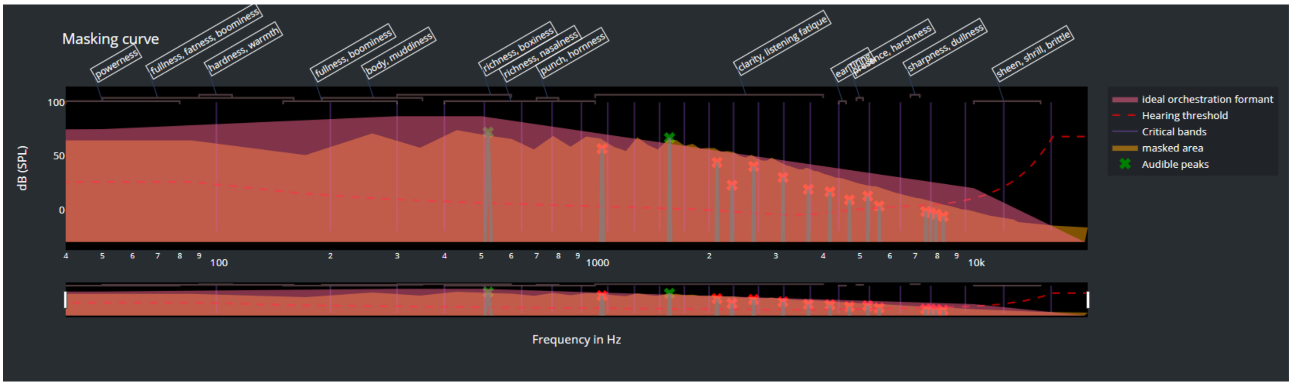
[240] Chen, Härdle, and Unwin 2007, p. 59.

[241] Chen, Härdle, and Unwin 2007, p. 60.

9.4 MASKING GRAPH VISUALIZATION CHOICES EXPLAINED

The first example is the main graph in the App’s Chord section. The graph provides a great deal of information in a condensed space, but my intention was to present the information in an intuitive way that is understandable to a musician.

Let’s examine the main graph, which shows the masking curve of the Chord section of the Score-Tool App in Example 37.

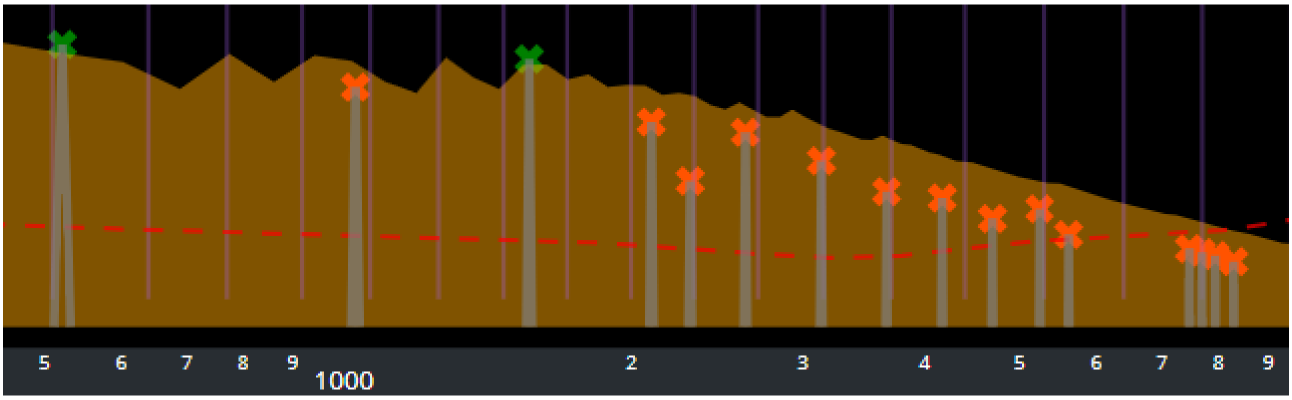


Example 37. Score-Tool App’s main graph represents the masking curve of orchestration input as a trombone, two trumpets and a flute, a grouping set as a target. All instruments play *mf*. The *x*-axis represents the frequency on a logarithmic scale, which increases from left to right, with the *y*-axis representing the sound pressure in dB.

This is the view the user sees when the input orchestration is a trombone, two trumpets, and a flute set as the target. The view is familiar to those who work with digital sound because it resembles a single window sound spectrum view. The difference between a traditional sound spectrum and the Score-Tool masking curve is similar to the difference between a musical score and a Schenkerian graph. The masking curve graph represents a distilled version of the spectrum, showing the aspects that are especially useful for a composer and orchestrator rather than showing the objective raw spectral data. As with a Schenkerian graph, only the visual design is the same with the raw spectrum. Some of the graph’s information is repeated from a different point of view in other graphs or turned at comparable values, but here I concentrate on explaining the possibilities for the visual exploration.

The first thing to notice is the impression of general audibility, or in other words, how the target instrument’s spectrum clashes with the orchestration. The clash is indicated by showing only the loudest peaks of the target, which represent the audible pitch with the overtones we would hear if we heard the target instrument when it plays the tone. The colored area indicates the threshold of masking in different critical bands in our hearing system, which is calculated from the orchestration data by putting it through the mathematical model.

If we zoom in to the area where the target peaks lie, we see how the graph lets us compare the power of individual components of the target sound to the masking effect of the orchestration, as seen in Example 38.

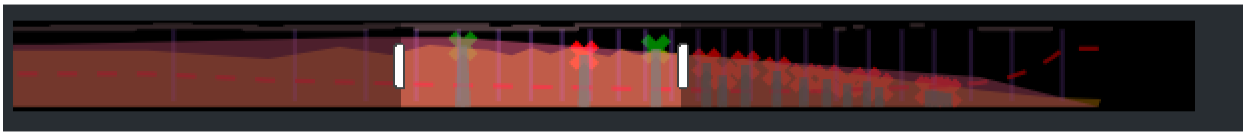


Example 38. Spectral peaks of the target instrument show in marked green above the masking curve and turn red if they fall below the masking curve, i.e., they become inaudible.

The visual aid provided is the neutral color of the masked area against the traffic-light coloring of the target peaks. The neutral color indicates that there is nothing good or bad in the masked area; it is just a phenomenon that happens when we hear the sound. As for the red- and green-colored peaks, red indicates the orchestration is contrary to the user's preference; if an instrument is selected as a target, I assume the user wants it to be audible. The grid of vertical lines with neutral purple coloring indicates an approximation of the borders of the critical bands of our hearing system. The bands are faint-colored because they are needed only in cases where multiple target peaks reside inside one critical band. In the case of a single target, this does not happen often, and when it does, the graph shows if peaks are close or far apart within the same band. In the case of noise components, all activity within one band is heard as a single component, but with tonal sounds (here meaning pitched), it is possible to hear peaks that are far apart as two distinct components. The coloring is the important part of the graph, since colors are the essential part of its design, while in practice (visual) color is one of the most difficult aspects to get right.[242]

While exploring the masking curve, one must keep in mind that the graph gives a monaural estimation, which can be altered simply by panning the sound sources in the horizontal hearing space. This was discussed for example in Part I, section 6.4. In other words, the graph does not give a definitive answer to the question, “Can I hear the target if I use this orchestration?” The role of the visual feedback in masking is to alert the artist what to be aware of in orchestrating. If the masking curve gives mainly a “green light” for the target, it gives the artist the freedom to experiment with unusual instrument combinations in the orchestration. If, however, the indication is on the “reddish” side, it might be wise to stick to the known conventions in that particular passage.

There is also a zooming panel below the spectral information, shown in Example 39.



Example 39. A close look at the zooming panel in the Score-Tool App.

---

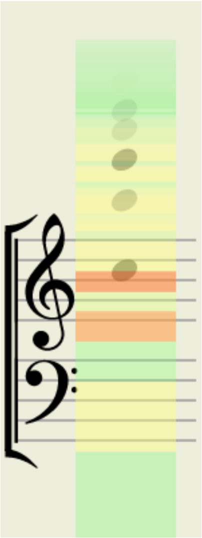
[242] Chen, Härdle, and Unwin 2007, p. 68.

The zoom feature gives an idea of where more detailed work could be done on a score. The argument might sound peculiar, but “zooming” into the timbre enables a composer to pick up features that might otherwise go undetected. This level of detail is essential for artistic work, which many times involves actions that may not be clearly audible in the final composition, not even to the composer. This feature might give a composer the courage to experiment more than usual with the parameters of orchestration.

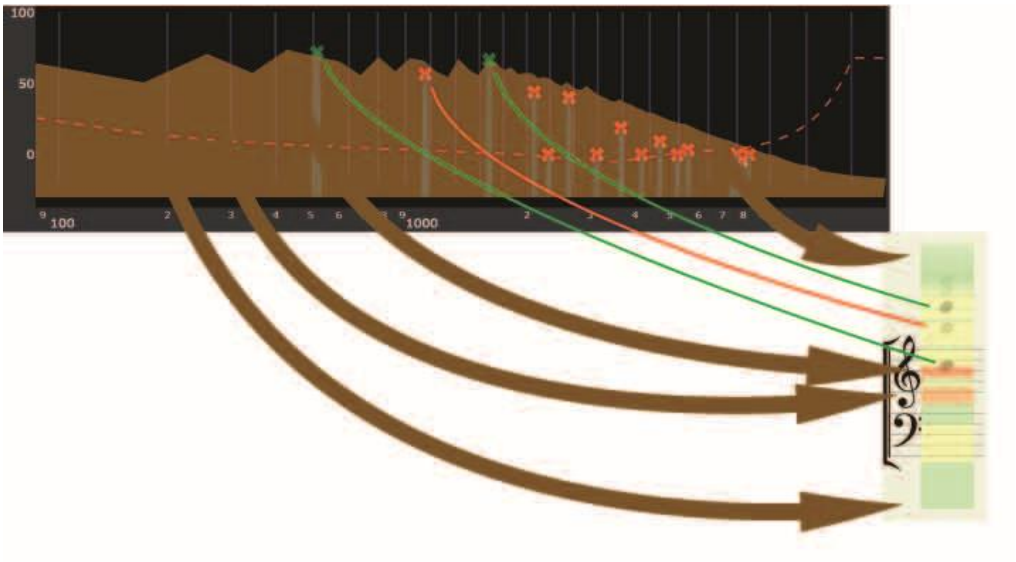
I now present another graph. It also shows the masking area and the target peaks, but in a different orientation and with a different background from the masking graph.

9.5 STAFF GRAPH – REDUNDANT INFORMATION

In the Score-Tool App, I included a graph with redundant information on masking and the target, information intended specifically for musicians. The staff in the score can be seen as a frequency space, where each note symbol represents a frequency equivalent to the fundamental of the notated pitch. The frequency space drawn on the staff is a representation closer to a musician’s experience than the frequency-decibel graph presented earlier. On a traditional staff, frequencies are represented vertically in an intuitive way, with low frequencies at the bottom and high frequencies at the top. Here, the corresponding decibels are marked with colors, where the color scale, staying true to the traffic light scheme, is green-yellow-red. The green indicates low, yellow indicates medium high, and red indicates high masking in the frequency area. The target peaks are marked with a black notehead, with a displacement in normal notation indicating the sharpness or flatness. For example, a notehead between *b* and *c* means a frequency corresponding to the fundamental of a microtone-sharp *b*. The faintness of the notehead indicates the power of the current peak. Example 40 shows an image of the frequency masking graph on a musical staff.

The audibility of the target can be estimated by the color: if there is reddish or a strong yellow color on a faint notehead, this indicates that the masking is strong on a weak partial.

Example 40. In the Score-Tool App, the masking curve is shown on a musical staff, where the heaviest masking is in red, the second heaviest in yellow, and “safe” registers in green. The faintness of the black notehead indicates the strength of the target partial. The audibility of the target can be estimated by the presence of a reddish or strong yellow color on a faint notehead. This indicates that the masking is strong on a weak partial.

In general, I recommend using the masking graph for exploring the orchestration data, because there is more information embedded there than in a staff graph. The main purpose of a staff graph is to help a musician understand the masking graph by presenting the data with musical symbols. The correspondence of graphs in Example 38 and Example 40 may be trivial yet it is shown in Example 41 for added clarity.



Example 41. The correspondence of the two masking graphs in the Score-Tool App.

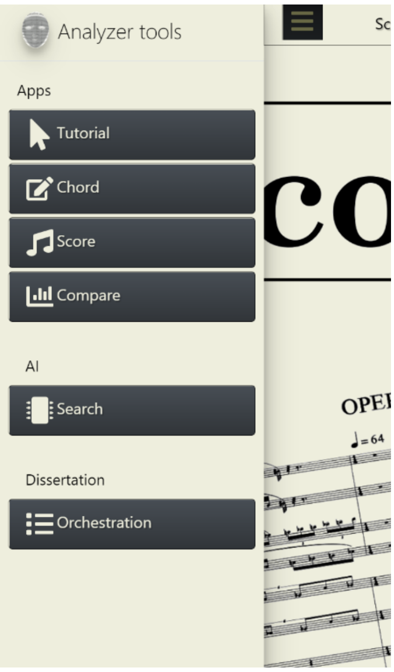
The corresponding masking areas are indicated with brown arrows and the corresponding target peaks are indicated with green and red lines. These arrows and lines are only shown in this picture; they are not included in the App.

10 FEATURES OF THE SCORE-TOOL APP

The purpose of the Score-Tool program is to test and analyze material that has already been produced; the program itself does not orchestrate or compose music. It is a tool for testing, pre-evaluating, and proof-reading purposes. All the components and libraries used are deliberately chosen so that the project is completely open source and can be distributed and modified according to individual needs.

The program is used in a browser window. The compatibility of the individual browser is not guaranteed, since the program uses some advanced html5 and css3 features that are not implemented on all browsers. In its development, I have used Google Chrome. Most of the program functions happen on the backend, so all calculations are done on the server side. The user device only renders the viewed pages.

The opening page shows the main menu on the left-hand side.



Example 42. The Score-Tool main menu

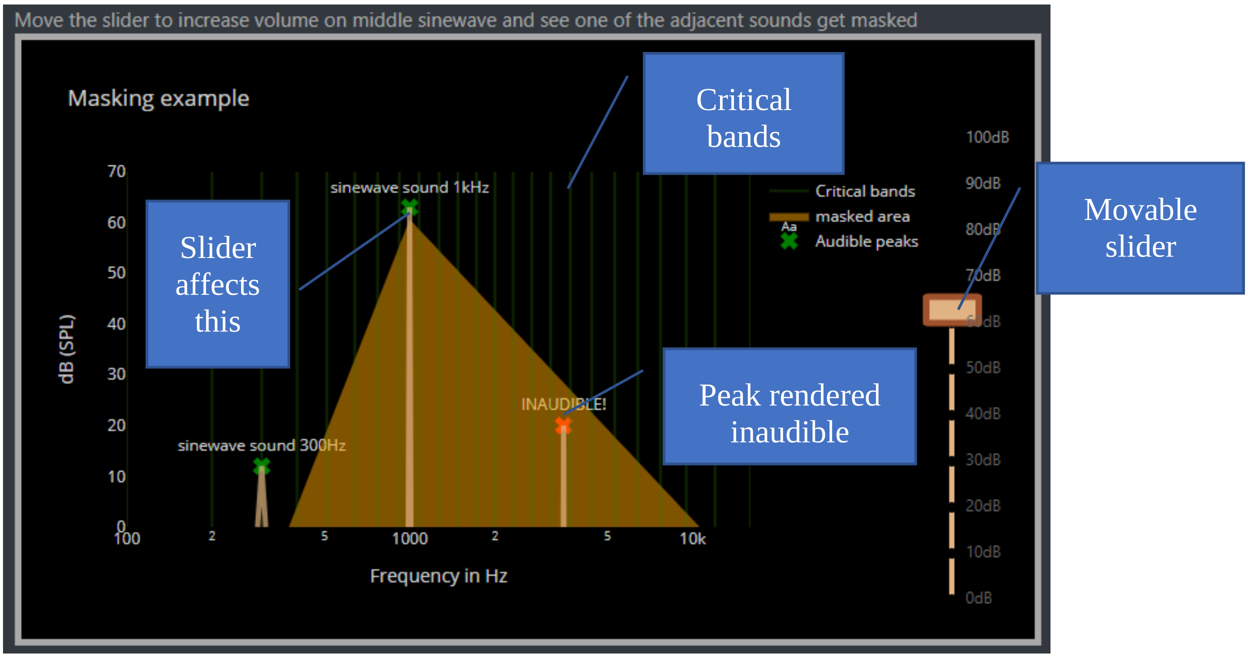
Example 43. The opening page of the Score-Tool App

The main menu consists of 4 sub-apps, a section for searches, and this project report as a hierarchy-tree. The sub-apps are called Tutorial, Chord, Score, and Compare. The major one is the Score App, which contains most of the code. The Tutorial and Compare Apps are for informational purposes, while the Chord App is for quickly checking an individual orchestration chord. The Search feature is under development; it is already usable, but is not discussed here.

10.1 TUTORIAL

In the Tutorial App, there are interactive demos showing the concepts used in the main part of the program in simple terms. In the masking example, the user can see how the amplitude of the sinewave component affects the masked frequency area. In the graphic chart, three sinewave peaks are marked, the static ones being on the left and the right-hand sides. The amplitude of the middle one can be adjusted with the slider on the far right. The critical bands are marked in faint green against the black background.

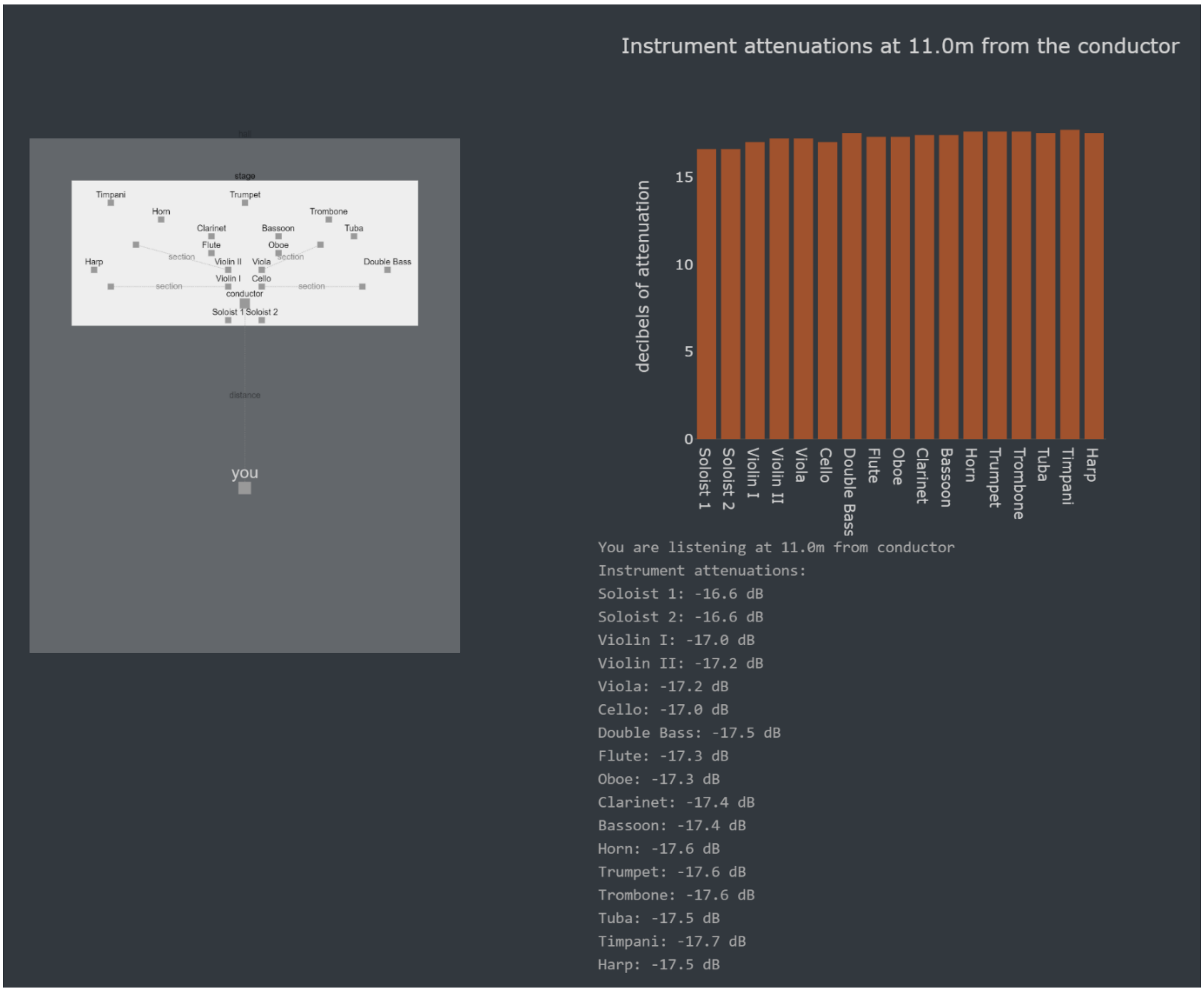
Moving the slider up shows that the masked area spreads more towards the high frequencies than towards the low frequencies. When the amplitude rises, the masking area eventually reaches the sinewave peak at 3.5 kHz and renders it inaudible. The chart is the same form used on the other apps, with a logarithmic frequency scale on a horizontal axis and the SPL dB scale on a vertical axis.



Move the slider to increase volume on middle sinewave and see one of the adjacent sounds get masked

Example 44. Tutorial on auditory masking. The slider affects 1 kHz sinewave component amplitude.

There is also a section in the Tutorial where the user can test the effect of seating placement on the loudness of orchestral instrument sounds. On the left-hand side, there is a figure with instrument names, a conductor, and You. The last one, "You," indicates the listening position. Each object can be dragged with the mouse, which changes the sound attenuation vis-à-via the listener (You). Try, for example, to drag "you" near the orchestra and see how that affects the balance of the orchestra; the instruments near you attenuate less, i.e., the sound is louder the closer it is to your position. The calculations are made for an average hall, with a critical distance of 7 meters (see Chapter 10, section 10.3 for details).

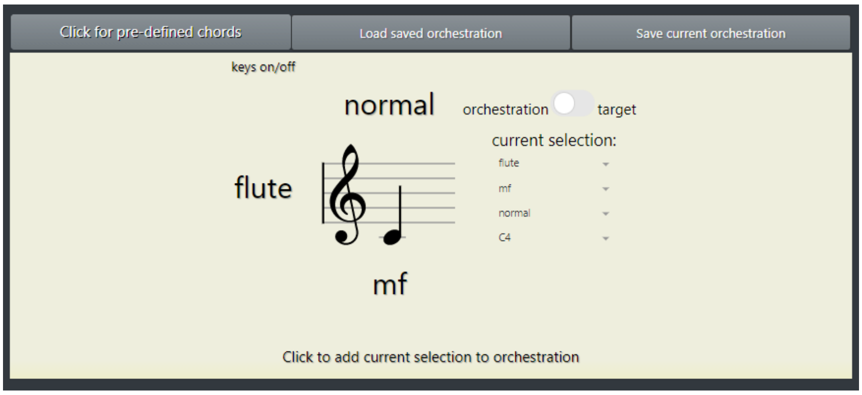


```
You are listening at 11.0m from conductor
Instrument attenuations:
Soloist 1: -16.6 dB
Soloist 2: -16.6 dB
Violin I: -17.0 dB
Violin II: -17.2 dB
Viola: -17.2 dB
Cello: -17.0 dB
Double Bass: -17.5 dB
Flute: -17.3 dB
Oboe: -17.3 dB
Clarinet: -17.4 dB
Bassoon: -17.4 dB
Horn: -17.6 dB
Trumpet: -17.6 dB
Trombone: -17.6 dB
Tuba: -17.5 dB
Timpani: -17.7 dB
Harp: -17.5 dB
```

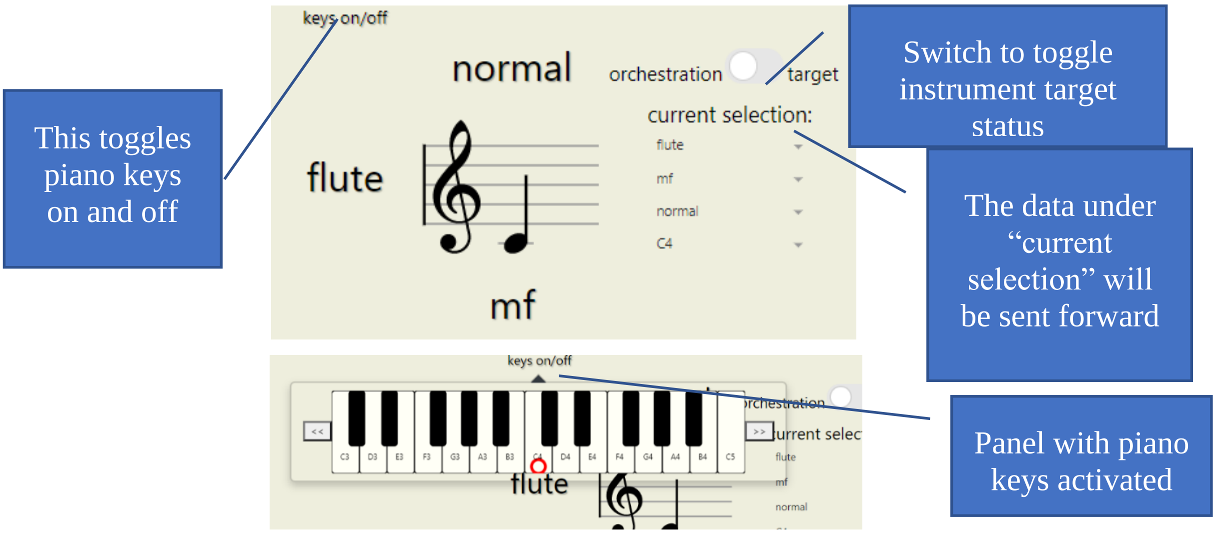
Example 45. The interactive graph shows the attenuation of instrument sounds in the hall. Click and drag the listener (You), or any instrument, to see the effect in attenuation.

10.2 CHORD

In the Chord App, you can compose an orchestral chord and test its masking and color properties. The Chord App is suitable for quickly prototyping an orchestration or for a rapid check of whether a soloist's sound will be audible. Opening the Chord App brings up a sub-window with a few buttons and a staff.



Example 46. The input panel of the chord app.



This toggles piano keys on and off

Switch to toggle instrument target status

The data under "current selection" will be sent forward

Panel with piano keys activated

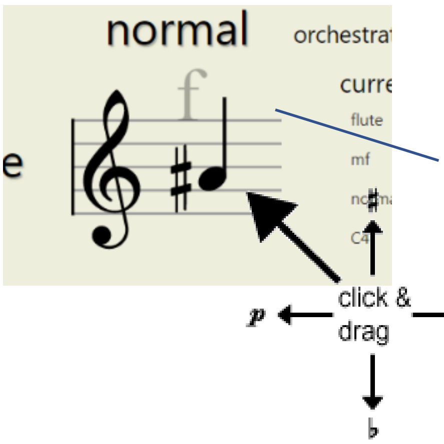
Example 47. Single instrument input panel in the Score-Tool Chord App.

The gray buttons at the top allow you to load and save orchestral chords. The pre-defined chords are for demonstration purposes: to show quickly how the App works without having to choose the instruments yourself. The Save button lets the user save an orchestration as a text file on the local machine. The text file has a structure that can also be edited with a text editor, but currently the load algorithm does not correct errors, so invalid files are not loaded. A Load button behaves the same way as the pre-defined chords button, only it asks the user to pick up a text file from their machine, preferably one that has been created earlier with the program. At the bottom, there is one more button: "Click to add current selection to orchestration." This button adds a user-selected instrument with properties to the chord. Instruments can also be added to pre-defined and user-loaded chords.

Adding an instrument can be done in the middle section in the sub-window. The important thing to remember is that the parameters under the text “Current selection” are sent to the Chord. Keep an eye on these parameters, which are updated, when you make modifications on the staff. At the upper left corner is a small button to toggle piano keys on or off if you want to input a pitch from a piano keyboard.

Clicking any of the parameters on the right side of the staff allows you modify them. Clicking an instrument name brings up a menu containing all the instruments in the database. Clicking “Technique” allows you to select all the techniques available for the current instrument, and clicking the dynamic marking lets you select the dynamics.

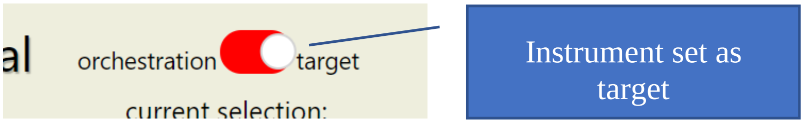
Clicking the staff itself selects the pitch, while clicking over or under the staff changes the clef. There is also a quick select feature: if you click and drag on the staff, dragging up selects a sharp, and dragging down selects a flat accidental. Dragging left selects *p*, and dragging right selects *f*.



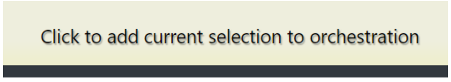
Click the note and drag up to make it sharp and drag right to make it f

Example 48. Instrument accidentals and dynamics can be edited by the mouse.

The important toggle is the one on the upper right, which lets you set the instrument as a part of the orchestration or as target. This selection can be changed later.



Example 50. Instrument target toggle enabled.



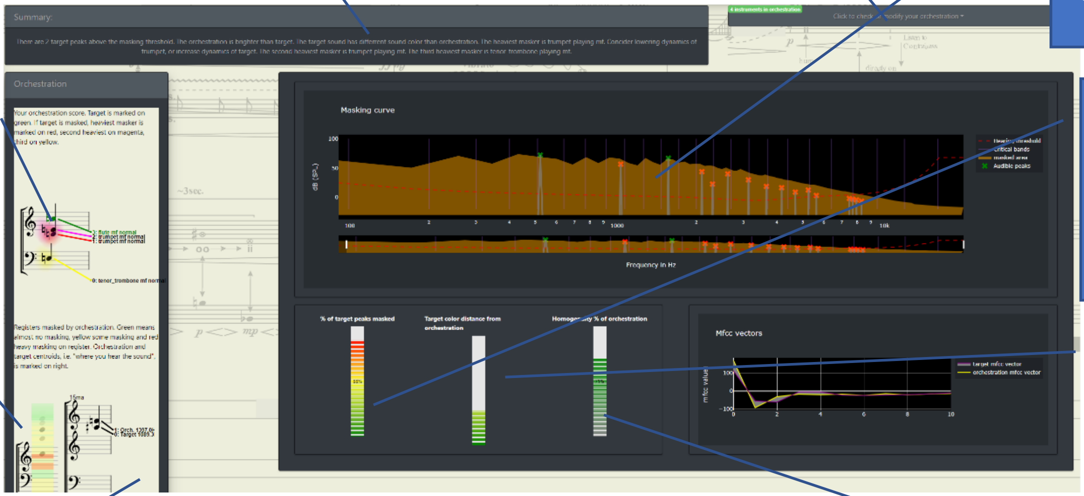
Example 49. Clicking this button adds the selection to the orchestration chord.

Once the parameters under “Current selection” are correct, they can be sent to the Chord by clicking the button at the bottom.

The update of the page may take a few seconds. The App is a “single page app,” so only the components needed are refreshed, not the whole page.

After a Refresh, the chord analysis section is loaded. This section consists of several sub-windows, showing the analysis data for the orchestration chord (see Example 51). For a quick check, there is a summary on the upper left, which describes verbally the masking and color relation between the target and the orchestration. In the summary, there is also an estimate of whether the target is audible. On the right in the summary there is a clickable menu where the current orchestration can be modified.

the masking graph are three percentage gauges that measure certain target parameters against the orchestration.



Summary:

There are 2 target peaks above the masking threshold. The orchestration is brighter than target. The target sound has different sound color than orchestration. The heaviest masker is trumpet playing mf. Consider lowering dynamics of trumpet, or increase dynamics of target. The second heaviest masker is trumpet playing mf. The third heaviest masker is tenor trombone playing mf.

Orchestration

Your orchestration score. Target is marked on green. If target is masked, heaviest masker is marked on red, second heaviest on magenta, third on yellow.

Registers masked by orchestration. Green means almost no masking, yellow some masking and red heavy masking on register. Orchestration and target centroids, i.e. "where you hear the sound", is marked on right.

Example 51. The analysis results of the orchestration chord.

Summary of the analysis in text

Button to modify orchestration

Masking curve and target partials view

The current orchestration with color-coded maskers

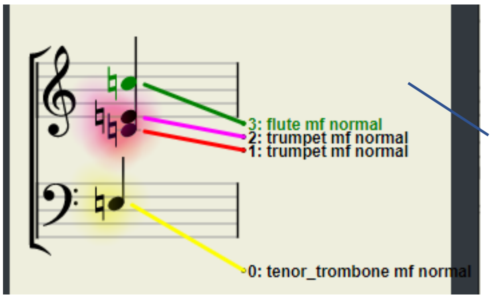
Percentage of target partials masked

Masking curve in musical staff

Target timbre distance from

Centroid view in musical staff

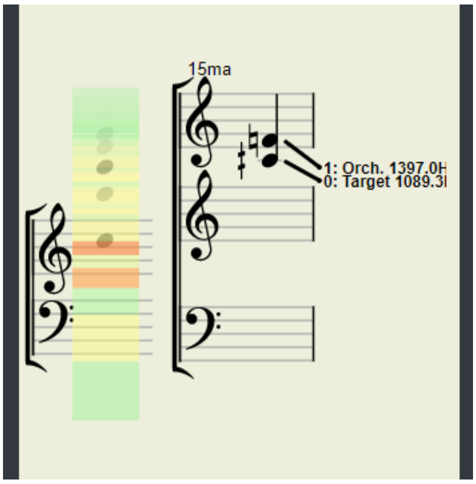
Homogeneity percentage of the orchestration



Clicking the instrument name toggles its target status on and off

Example 52. The orchestration view is color-coded. Red indicates the strongest masker, magenta the second strongest masker, and yellow the third strongest. Green indicates the target. Clicking the instrument name toggles the status between target and orchestration.

Above the box with musical staves, there is a staff system showing the current orchestration on the bass and treble clefs (see Example 52). The noteheads of the orchestration are black, and the noteheads of the targets are green. The notehead of the heaviest orchestration is shaded



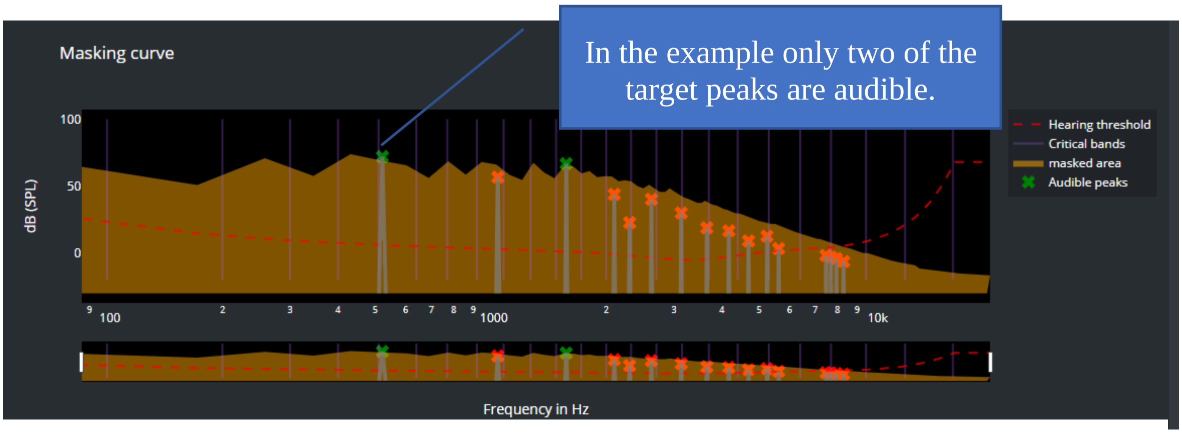
Example 53. On the left-hand side, the masking curve is plotted on a musical staff, which is the "musician's view" of the masking graph. Red means heavy masking, yellow moderate, and green little or no masking. The target peak strength is marked with faint black noteheads. On the right-hand side, the spectral centroids of the target and orchestration are plotted on a musical staff along with the exact centroid Hertz rating.

with red and follows the red line. The second heaviest masker is shaded with magenta and the third with yellow. The instrument names, dynamics, and techniques are marked on the right. Clicking an instrument name toggles that instrument as the target or as an orchestrational component. The effect is calculated immediately after the click and may take a few seconds.

Under the color-coded orchestration are two staff systems (see Example 53). On the leftmost system, the excitation on critical bands is shown on the staff. The bands that are heavily masked are colored red, the intermediate masks are yellow, and the light masks are green. The target partials are marked on black noteheads. The transparency tells the amplitude of the current partial. The black noteheads show no accidentals, but they are marked a bit over their usual place to indicate the sharpness or flatness. Thus, a faint notehead under a red critical band means low audibility, and a solid notehead at the top of a green band means good audibility.

The staff on the right-hand side shows the spectral centroid on the notated staff. After the instrument's name, the centroid frequency is marked on the Hertz scale. Here, if the target centroid is larger than the orchestral centroid, it means that the target sound is somewhat brighter than the orchestration sound. A bright sound is generally more audible than a dark sound. In Example 53, the target centroid is lower than the orchestration, and the target's overtones are located around the red and yellow areas, so according to the graphs, the target's audibility is poor.

The masking curve graph shows a detailed view of the target partials vis-à-vis the orchestration masking curve. The masking curve is the same as that on the color-coded staff in the previous example, but here it is plotted in decibels of the function of frequency. The critical bands are marked in faint purple against the background. The human hearing threshold is marked on a dashed red line. Anything under the dashed red line is undetectable by the ear. Audible peaks come from target instruments. Each peak represents each partial of the target. For clarification, if the peak is green, it is above the masking threshold; if the peak is red, it is below the hearing threshold. Under the main graph the same graph appears with white vertical lines on each end. By using the carets, the user can zoom in on the graph and easily find a place of special interest. Double clicking the graph resets the view to the original. The graph is interpreted by checking the audible target peaks on each critical band. Placing the mouse cursor over the peak shows the peak's frequency and amplitude. If the peak is green and clearly above the masking and hearing levels, then it is probably audible. However, if there are just a few peaks of the target above the threshold, as in this example, then the instrument might no longer be audible, since most of the peaks are under the threshold.



In the example only two of the target peaks are audible.

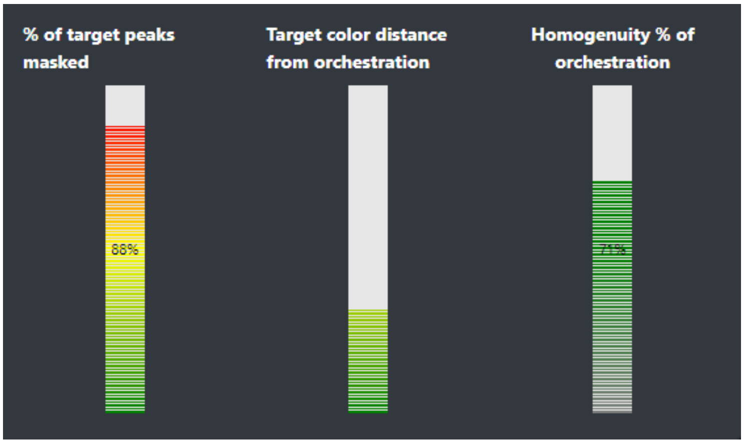
Example 54. The main masking graph. Masking caused by orchestration is in the light brown area. Target peaks appear in red if they are inaudible and in green if they are audible. The dashed line indicates the hearing threshold, and the faint purple lines mark the critical bands.

The three gauges under the masking graph show simplified information from the algorithm's results. The first gauge shows the percentage of target peaks under the masking threshold compared to all target peaks. It does not give a direct estimate of audibility, because there is no information about the target's amplitude. This can be considered a rough masking estimate.

The second gauge measures the target color distance to orchestration by showing the Euclidian distance of the target and orchestral MFCC vectors. The full bar means the target timbre is far from the orchestration timbre. Values under half a bar can be considered roughly close to the colored sounds.

The third gauge shows the value of the coefficient of variation algorithm normalized as percentage values. The homogeneity percent of orchestration does not count the target instrument(s) from the orchestration. If the value is high, the orchestration is homogeneous, as is the case in a string quartet, for example. This means that if at the same time the target color distance is high, the target might be audible despite a high percentage of masked peaks. However, my intuition says that if the orchestration homogeneity is low, even a target in a distant color could be hard to detect, since the variety of orchestral colors is already high. This would need to be tested empirically.

The gray sub-window button in the middle-right opens a drop-down menu where the orchestration can be modified (see Example 56). A green badge at the top of the button shows the number of instruments currently entered. When the menu is activated, the background of each instrument slot is color-coded in the same fashion as on the orchestration staff, namely, the target is green, the heaviest masker is red, the second heaviest masker is magenta, and the third is yellow. There are quick prototype buttons of octave up-and-down transpositions for testing the effect of changing the register without otherwise affecting the harmony. Here the individual instruments can also be switched on and off with the small toggle (another toggle switches the target status on/off). The general usage would be to test a change of register or the dynamics of the target or the heaviest masker. A change in every parameter can be done here, but some things, such as a change of pitch or instrument, can be more easily done on the staff input window. Changing any parameter here affects the graphs and color-coding in real-time. If the change results in complex re-calculations, the effect of the change may take a few seconds. Thus, rapid change of parameters in sequence is not recommended.



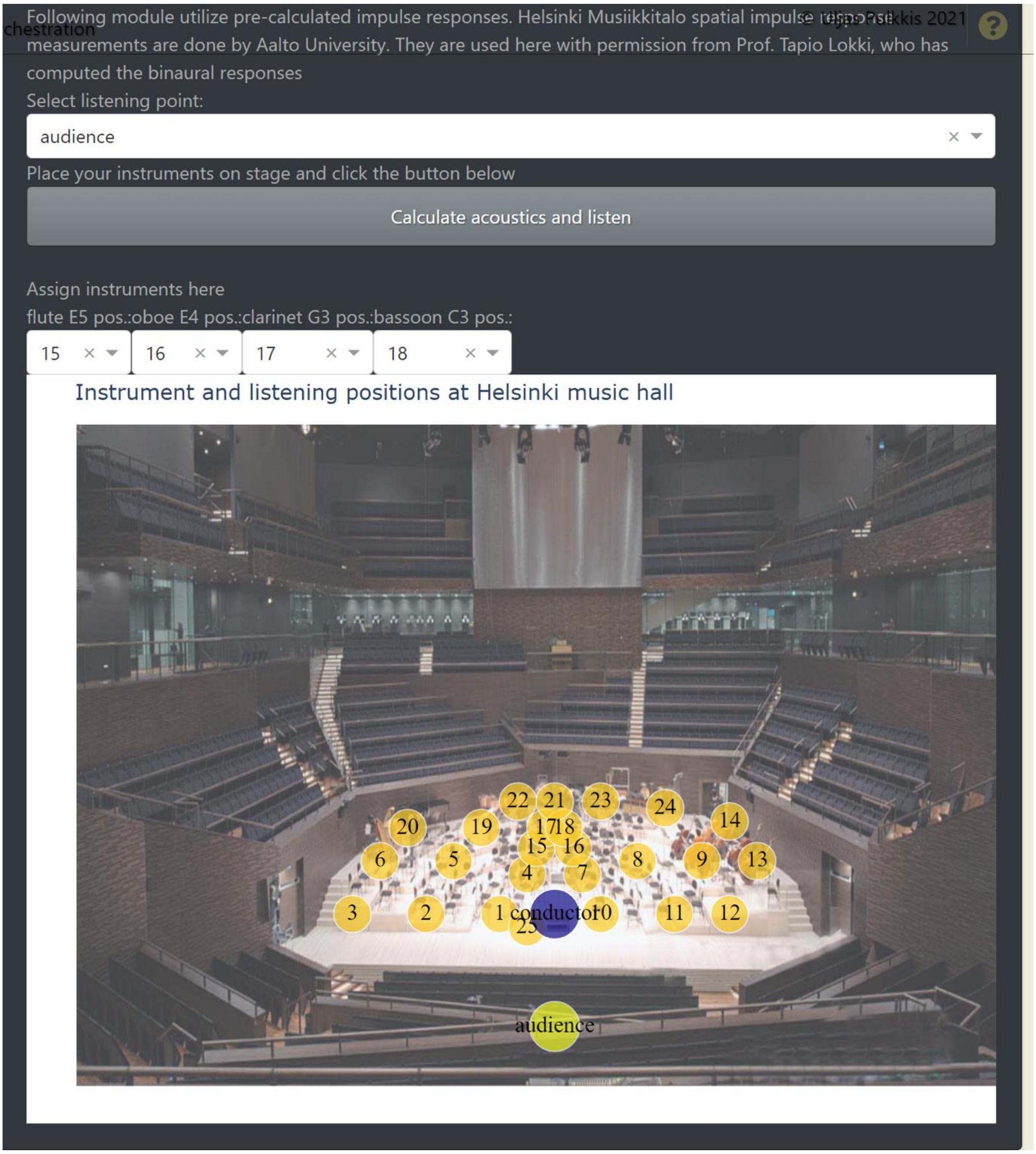
Example 55. Three gauges view that show (from left to right) the percentage of target peaks masked by orchestration, the euclidian distance of target mfcc vector to the orchestration mfcc vector, and homogeneity, i.e., the coefficient of variation of the individual instruments in the orchestration.

Example 56. In the orchestration modification panel, the background of the instrument slot is red if the instrument is the heaviest masker, magenta if it is the second heaviest, and yellow if the third. The target background is green.

Interesting orchestration chords, i.e., chords worth saving, can be saved at any time on the local machine by clicking the “Save current orchestration” button. However, any change in the Chord App remains intact, even if the user changes the application (such as to the Tutorial and back) as long as the page is not refreshed. Refreshing the page means restarting the App (as is the case with all web apps).

It is also possible to listen to the current orchestration sound with an implemented simulation of the main concert hall in Helsinki’s Music Centre. On Score-Tool, you can assign each instrument a pre-defined place on stage and choose a listening position, either from the audience’s perspective or at the conductor’s podium. The audience’s seat is on row 6 in front of the orchestra, beyond the critical distance of the hall.

Clicking the button “Calculate acoustics and listen” renders the orchestration in three versions: target only, orchestration only, and target with orchestration.



Following module utilize pre-calculated impulse responses. Helsinki Musiikkitalo spatial impulse response measurements are done by Aalto University. They are used here with permission from Prof. Tapio Lokki, who has computed the binaural responses

Select listening point:

audience

Place your instruments on stage and click the button below

Assign instruments here

flute E5 pos.:oboe E4 pos.:clarinet G3 pos.:bassoon C3 pos.:

15 16 17 18

Instrument and listening positions at Helsinki music hall

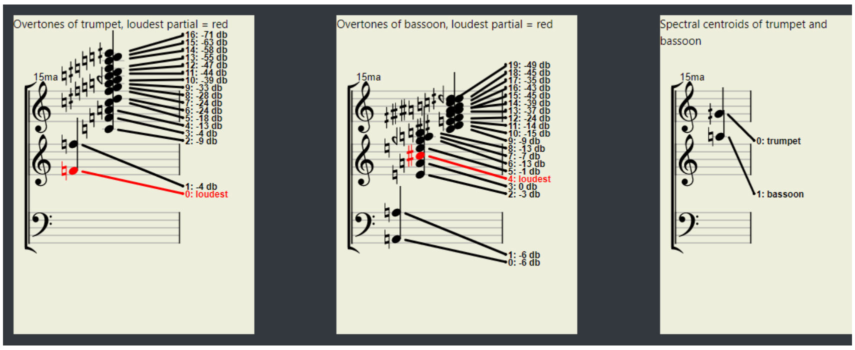
Example 57. Acoustic model of the main concert hall in the Music Centre in Helsinki. On Score-Tool, instruments can be assigned stage positions and a listening position can be selected either at the conductor’s podium or as an audience seat (row 6 in the middle). Clicking “Calculate acoustics and listen” renders a binaural simulation of the user’s orchestration.

10.3 COMPARE

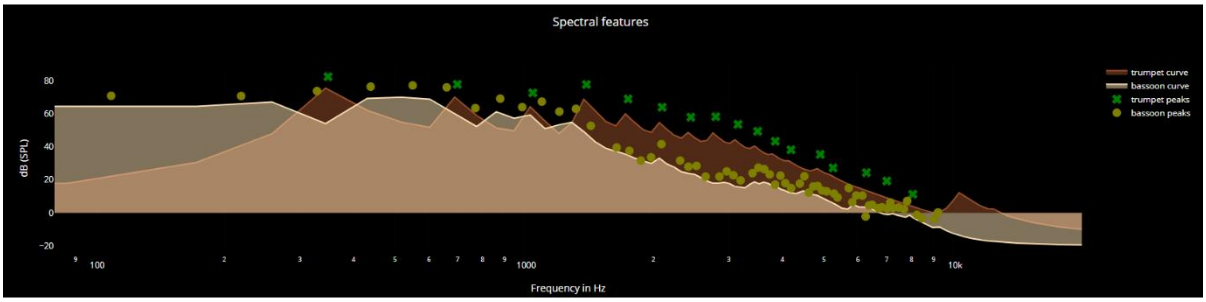
The Compare App enables a user to compare the properties of two instruments in the orchestral instrument database. The instrument selection has the same functions as the Chord App, and the instructions can be read in the previous section.

Clicking the button “Click to compare” brings up a sub-window with data placed on staves and graphs. The first two staves show the partials of chosen instruments on a musical staff. The frequencies of the partials are rounded to the closest quarter tone, and the tuning fluctuations may result from showing the root tone of a low-tone instrument as a microtone. On the right-hand side of the noteheads, the relative amplitude of the partials is listed. The loudest partial is shown in red and as 0 dB, and the loudness of the remaining partials is marked in relation to the loudest one. In this example, you can observe that the loudest partial of the low trumpet is the root tone, but the loudest partial of the low bassoon is, somewhat surprisingly, the fourth partial (the middle staff in Example 58). The third staff graph shows the centroids of the two chosen instruments in the same way as in the Chord App.

The spectral features graph in Example 59 shows the spectral peaks of both the chosen instruments and their masking curves. On this graph, the spectral content of the sound can be seen at a glance. The graph shows that even though both instruments have a rich spectrum, the trumpet sound has higher amplitudes in the sensitive hearing area, around 1-4 kHz. However, the low frequencies are also quite distinguishable by our hearing system, and the bassoon has a total of three partials below the trumpet root tone. According to the graph, the sounds of these two instruments have very different characteristics.



Example 58. Compare the App subwindow with the staves and the graphs with data. The staff on the left-hand side shows the partials of the first instrument as pitches on a musical staff. The middle graph is for the second instrument, and the right-side staff shows the spectral centroids of the instruments compared.



Example 59. The Compare App, another view of the data from the same instruments as in Example 58. The graph shows both the masking and the partial structure of both instruments under comparison.

10.4 SCORE

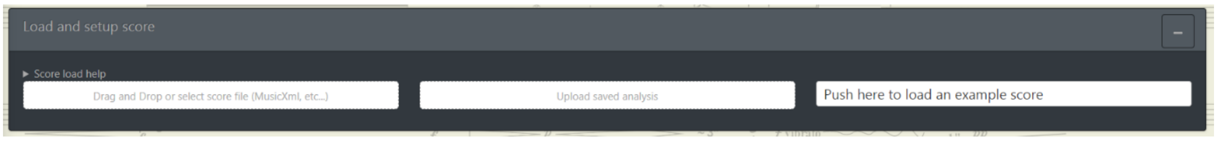
The Score App is the essence of Score-Tool. This App allows the user to analyze an entire score or a section of it with the algorithms described earlier in the report. The Score App is also computationally heavy, and for large scores with hundreds of bars, I recommended to run the program locally (i.e., run the server part of the App on a local machine). The calculated estimate of how long the analysis of the score takes does not include the server load, and multiple users analyzing multiple large-scale scores at the same time may result in extremely slow performance for each user.

The Score App opens with empty graphs and three white buttons at the top of the page (see Example 60). The three buttons let the user upload a score file, upload a previously saved analysis, or select an example score which is implemented in the program.

The button on the left-hand side accepts files which are dragged-and-dropped with the button. The button may also be clicked to open the file path. File formats that the program can read are all open-source formats, including those which major notation programs can export.[243] Regardless of the file format, the dynamics are flattened to three levels: *p*, *mf*, and *f*. I chose this approach to emphasize the directional nature of the masking analysis; the purpose of the program is not to fine-tune the dynamic of each player, but rather to give a rough estimate of the functionality of the orchestration used.

The middle button allows the user to upload an analysis text file saved with the program earlier. This makes it possible to bypass a time-consuming calculation made earlier and skip directly to the analysis data. Saving the analysis is highly recommended owing to the long calculation times.

The button on the right-hand side brings up a drop-down menu with example scores. Among other things, there is a "Test score," which I have created entirely for demonstration purposes as well as the first part of Brahms's violin concerto (Only in old.score-tool.com) for testing analysis on a large-scale score. In trying the program for the first time, I recommend trying the "Test score" and experiment with it to find out more about the program's features.



Example 60. Three options for selecting a score: uploading your own score, uploading a previously saved analysis, or selecting an example score.

After a score is selected for analysis, a menu appears from where the user can map the score's instruments on the analysis samples in the database (see Example 61). The table of parameters shows the names of the score's instruments under the heading "Score name." These names are read from the score file as they appear on every staff. The order should be the same as in a notation program; if the same instrument name appears multiple times, the order can be checked in the original score. Under the heading "Database name," the user can select the

---

[243] If you upload a midi file, make sure the dynamics are exported with proper velocity data because that is where the program reads dynamics.

database instrument which is the closest match to the notated instrument. The program tries to guess the choices, but I recommend hand-checking each one to avoid errors.

Under the heading “Technique,” the user can determine if the instrument calculation takes the special playing technique into account. As used here, “technique” applies to the whole section analyzed. Currently, I have not implemented an algorithm to retrieve a changing technique from the score. The choices for dynamics include the three-step levels (*p*, *mf*, *f*) along with the option “from score,” which results in applying the dynamic markings from the file.

The most important choice is under the heading “target/orch.” Here the user can select which instruments are counted as targets and which as orchestration. If the score under analysis is a solo concerto, for example, the soloist would be the natural choice for the target. Selecting multiple targets is possible, but the relevance for selecting half the instruments in a score as targets is very low. The last selectable parameter is to turn the instrument on or off for calculation. Among other things, this can be used to check how turning off a single loud instrument can affect the result. It can also be used to disable instruments not currently found in the database.

The light-brown double slider can be used to narrow down the number of measures for analysis. The default is the whole score, but the range can be narrowed to a minimum of one measure. The selected range is shown in brown text under the double slider and is updated automatically by moving each end of the slider.

To start the analysis, click the window-wide block button that also contains the estimate of the analysis time.[244]

---

[244] The time estimate is calculated for a gen 6 core i7 laptop processor on a pc.

Instruments as they appear in the score

Instruments in the database that match the names in the score

Either use dynamics from the score or set the fixed value

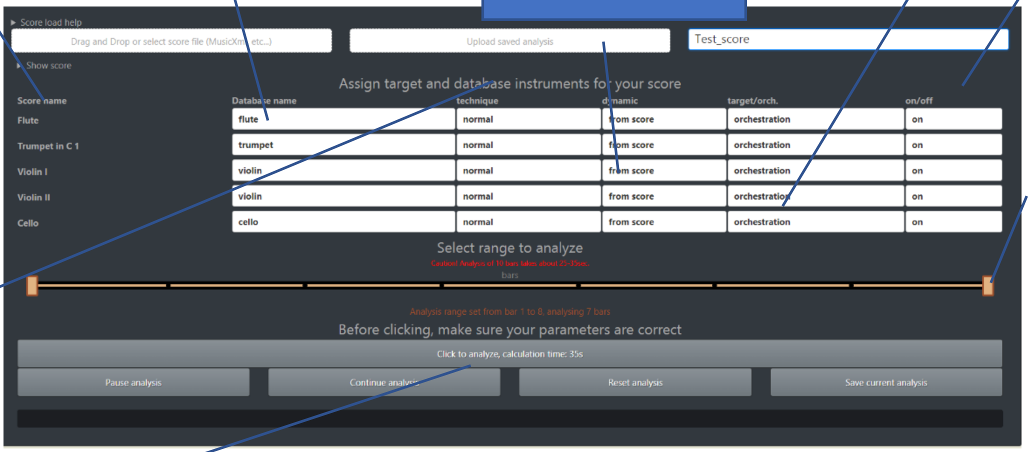
Select the staff line role: target or orchestration

Toggle staff on or off

Select fixed playing technique for staff

Select analysis range in bars by clicking and dragging the two slider handles

Click here to start the analysis



Test_score

Assign target and database instruments for your score

Score name	Database name	technique	dynamic	target/orch.	on/off
Flute	flute	normal	from score	orchestration	on
Trumpet in C 1	trumpet	normal	from score	orchestration	on
Violin I	violin	normal	from score	orchestration	on
Violin II	violin	normal	from score	orchestration	on
Cello	cello	normal	from score	orchestration	on

Select range to analyze

Caution! Analysis of 10 bars takes about 25-35sec.

bars

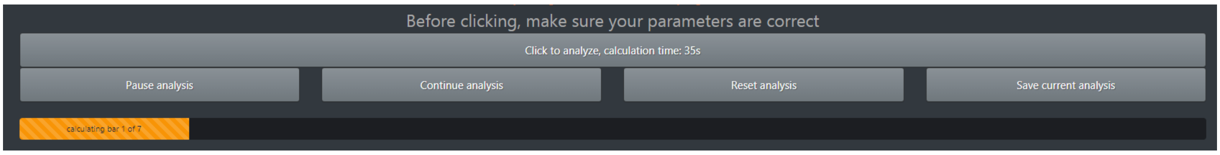
Analysis range set from bar 1 to 8, analysing 7 bars

Before clicking, make sure your parameters are correct

Click to analyze, calculation time: 35s

Example 61. Panel available after selecting a score.

The progress of the analysis is shown on the bar with the text indicating how many of the total measures are analyzed. The graphs are updated after each measure analyzed. The progress can be paused using the “Pause analysis” button and continued with the “Continue analysis” button. The “Reset analysis” button resets the whole page for a new score. Currently, because of an unresolved glitch, the reset button works only for an ongoing analysis, but not when the App is in the Pause state. The “Save current analysis” button can be pressed even when paused, but I recommend waiting until the task is finished.



Before clicking, make sure your parameters are correct

Click to analyze, calculation time: 35s

calculating bar 1 of 7

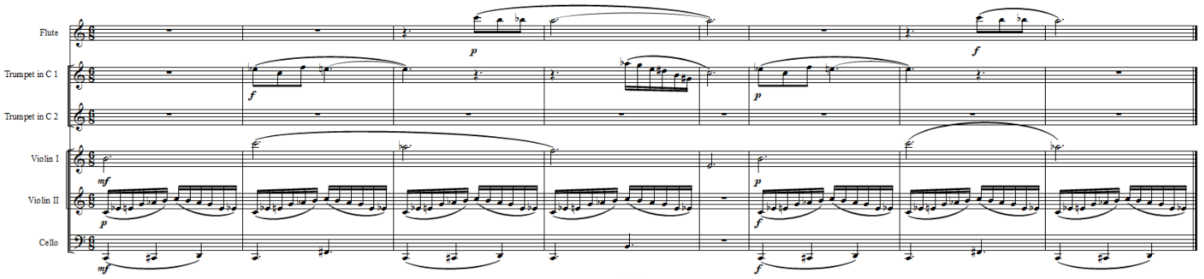
Example 62. Detail of the buttons that appear in Example 61. After clicking the analyze button, the progress bar appears showing how many bars are analyzed.

The resolution of the analysis is 0.1 seconds taken from the tempo of the score file.[245] The resolution is set according to the normal reverberation time in concert halls and the human temporal discrimination limit. In a dry hall, the resolution is probably smaller; in the acoustical setting of a church, it is probably larger.

---

[245] This is for the old.score-tool.com, in the new version there is no minimum resolution.

After the analysis is ready, you should see the analyzed score with the graphs rendered below. All the graphs have the same form, with measure numbers running from left to right on the *x*-axis, and the value changes on the *y*-axis (as seen, for example, in Example 65). All the analysis graphs are hoverable and clickable. Hovering shows the exact measure and the analysis value corresponding to the measure, and clicking the graph brings up a detailed chord analysis at the current point. As in the Chord App, the graphs show somewhat overlapping information both in a “musician friendly” and in a “scientific” manner.

The following graph examples are taken from an analysis of the “test score” that can be selected in the Score App start menu. The test score is a short 8-bar example I composed to demonstrate the use of the program. The instrumentation in the test score is flute, trumpet, two violins, and a cello. In the following examples, the flute has been chosen as the target instrument. The extra trumpet staff was added for debugging purposes to see how the program handles completely empty staves (they are ignored).

©UP

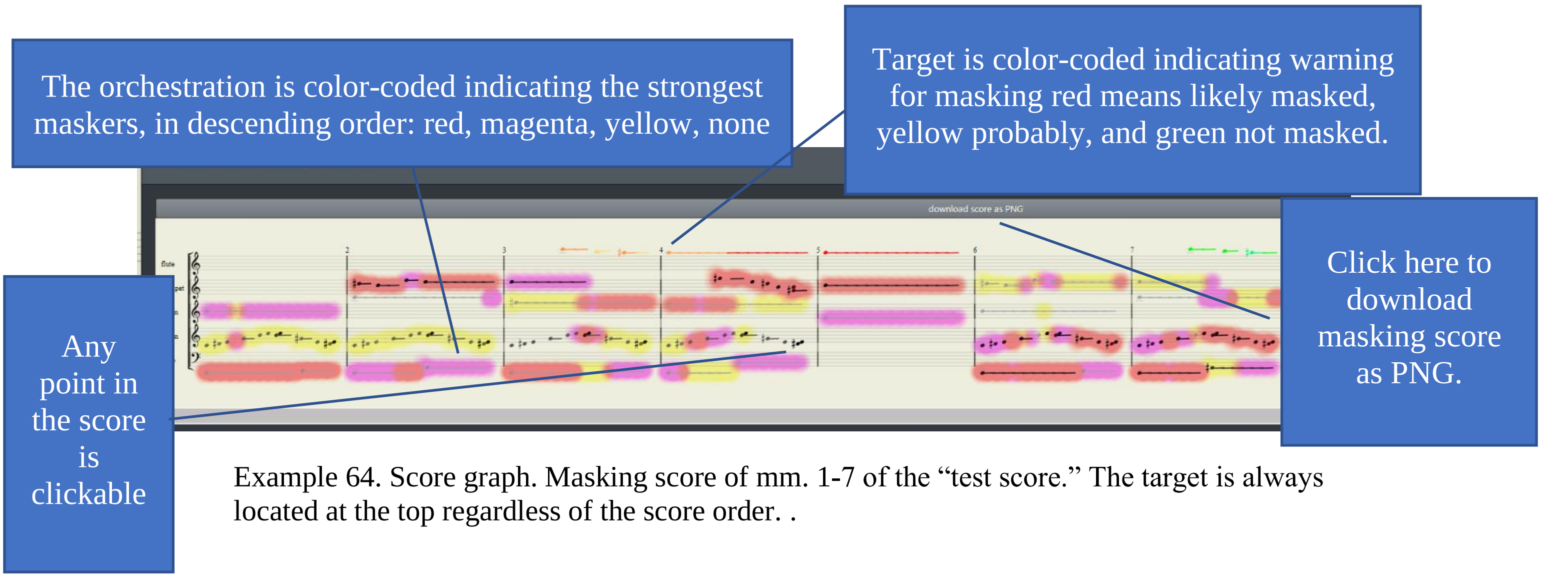
Example 63. The “test score” implemented in the Score-Tool App to provide an easy way to test the application’s functionalities. Try, for example, to set the flute staff as the target and see the results.

The first graph resembles a traditional orchestral score (see Example 64). The program places all target instruments at the top of the score, and the orchestration instruments in score order below. All measures have the same length, and the notated pitches are marked with a notehead in 0.1-second intervals. Natural signs for accidentals are not marked, even when they follow an accidental in the same measure. If the same note is sustained to the next time-interval unchanged, it is marked as a line, thus avoiding repeating the accidentals. Measure numbers are placed at the beginning of a measure at the top, and the score is horizontally scrollable.

The score notation is color-coded throughout, even in sections without the target, as seen in Example 64, where the target instrument has rests in the two first bars. Orchestration instruments are coded with colored notehead backgrounds, and target instruments appear as colored noteheads. The color-coding is the same as in the Chord App: the heaviest masker is marked in red, the second in magenta, and the third in yellow. If there is no target present, then the colors indicate the instrument with the loudest orchestration. The dynamics of the instruments are marked in transparencies, with *forte* noteheads 100%, *mezzo-forte* 75%, and *piano* 50% solid.

The target noteheads show the amount of masking at the current time-interval. A red target notehead means fully or almost fully masked. Magenta means under 80% of the partials, and yellow means that less than 70% of the partials are masked. Green coloring means that the target is probably audible. Target masking together with orchestration color coding already tells a good deal about the audibility of the target. The masking prediction can be verified by using the information from other analytical parameters in the graphs below or by clicking any notehead in the score. Hovering the mouse over noteheads highlights the clickable area, and the current clickable time stamp is shown at the top-left of the score sub-window.

The whole color-coded analysis score can be saved locally by clicking the text "Download score as PNG" above the score. This saves the entire score, even the parts that are not visible as a PNG image. The PNG image retains the transparency of color-coding and noteheads. If converted to, say JPG, the transparent colors turn solid.



The orchestration is color-coded indicating the strongest maskers, in descending order: red, magenta, yellow, none

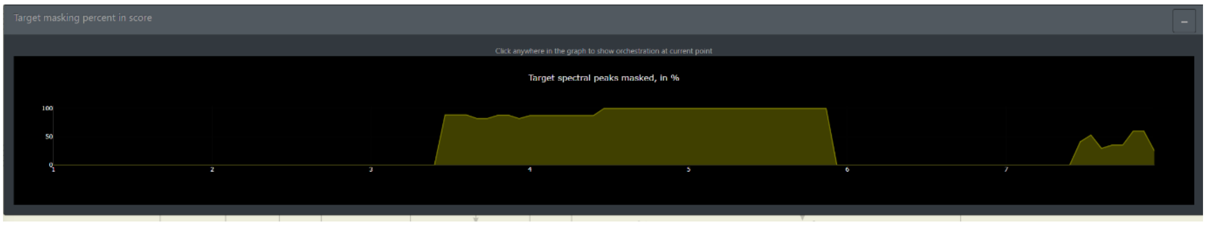
Target is color-coded indicating warning for masking red means likely masked, yellow probably, and green not masked.

Any point in the score is clickable

Click here to download masking score as PNG.

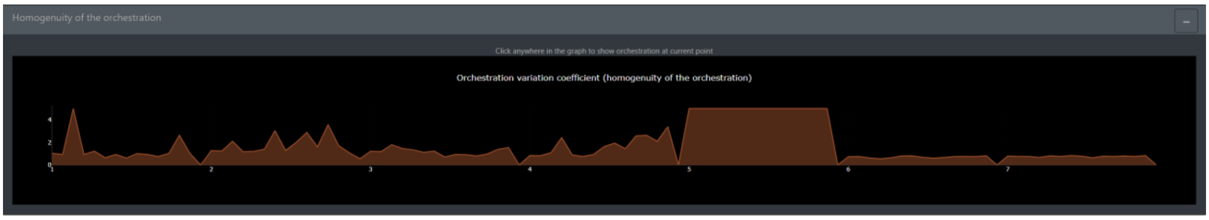
Example 64. Score graph. Masking score of mm. 1-7 of the "test score." The target is always located at the top regardless of the score order. .

The next graph (see Example 65) is the temporal representation of the masking percentage. The graph can be thought of as a detailed view of the color-coding of the target noteheads. The graph consistently has the same width, so the masking percentage is not visually comparable to the score graph, but measure numbers are marked on a horizontal axis. To check the details of a particular point, check the same point with the measure number on the score graph.



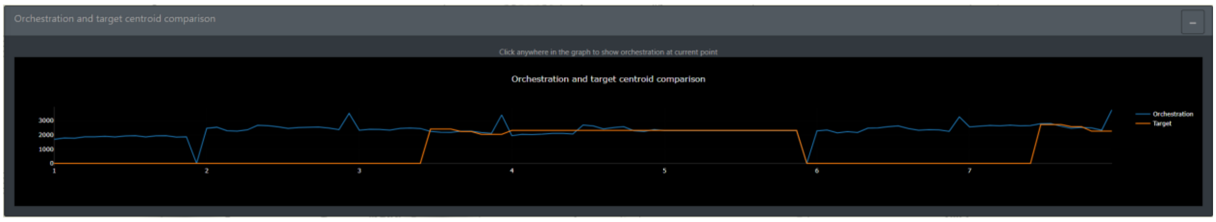
Example 65. Masking percentage graph. The graph shows the masking percent of the target as a percentage of target partials masked in the function of measure numbers. In this graph we see, for example, the target instrument (flute) entering the middle of m. 3 almost fully masked by the orchestration.

Below the masking percent graph is the variation coefficient graph (see Example 66). The values for this graph are drawn only from instruments marked as orchestration. The graph can be read as the homogeneity of the orchestration. The importance of a homogeneity reading increases when the orchestration is thick. This can be seen in m. 5 of the graph, where the orchestration consists just of two instruments with dissimilar sound colors, trumpet and violin; the variation coefficient jumps to the maximum.



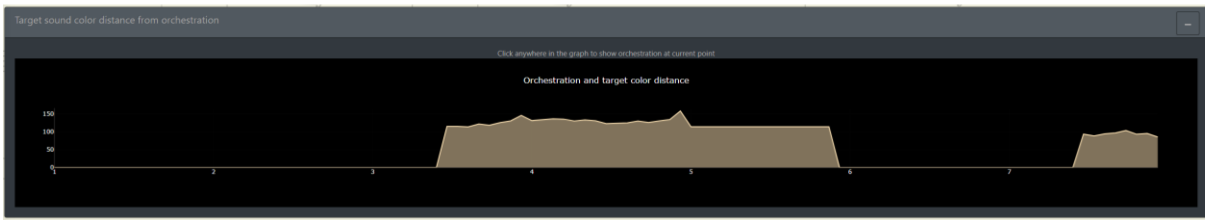
Example 66. Timbre homogeneity graph. A graph showing orchestration timbre homogeneity in the function of measure numbers. The higher value means the timbre is more homogeneous.

The centroid comparison can be read in the value “Is the target brighter than orchestration” in the graph shown in Example 67. The orchestration is marked in blue and the target in yellow. Sections without readings are marked 0. As in the previous graph, the *x*-axis is the measure number. The *y*-axis is the centroid value in Hertz. The definition of the centroid is explained in its own chapter.



Example 67. Spectral centroid graph. A graph showing the spectral centroid comparison between the target and orchestration. The orchestration is marked in blue and the target in yellow. When the target is not playing the value is 0.

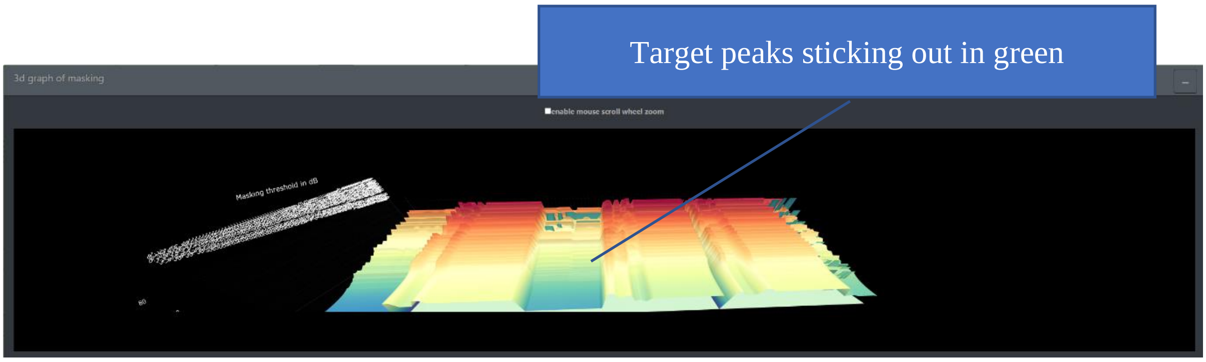
The timbre distance graph (see Example 68) gives the values of how distant the target’s MFCC vector is from the orchestration’s MFCC vector. The distance is calculated as a Euclidian distance. The *y*-axis value does not have a specific unit, but very distant timbres may be audible despite the yellow or orange color-coding on the score graph.



Example 68. Timbre distance graph. A graph showing the target timbre distance from orchestration. The value is a Euclidian distance of the MFCC vectors of the two.

The last score graph is a 3D representation of the masking graph from the Chord App. As in previous graphs, the horizontal value is the measure number. The *y*-axis shows the split critical bands as a bottom frequency of each band. The axis, or the “height of the mountains,” shows the degree of masking in decibels. The height is also indicated in color, from low values marked in blue, middle values in yellow and high in red. The target is marked in the same graph with green peaks. If a green peak sticks out from the orchestration, it means that the target sound is audible. Hovering the mouse over the graph gives the readings of both target and orchestration at the current time-point and in the current band.

The 3D graph can be navigated by rolling around its own axis dragging the mouse. The mouse wheel zooms the graph in and out. As the scroll-zoom can distract the scrolling the page, this feature can be turned off by clicking the box above the graph. Double click on the graph resets the view to default. A single click without drag brings up the detailed analysis window of the current time stamp.



Example 69. 3D graph. A graph viewing the masking graph of the orchestration in a third plane. The target peaks are sticking out as green peaks. The graph is zoomable and rotatable by clicking and dragging the mouse. The white element on the left is the frequency scale with low values in front and high values in back. Frequency values can be read either by zooming in or by hovering the mouse over the graph.

A click on any graph brings up a window with the same analytical data, just as in the Chord App, but here the chord is taken from the clicked timestamp. The clicked chord can be modified in the same way as described in the Chord section, but currently the modified chord is not transferred into the score graphs. As stated before, saving the analyzed score for later reloading is highly recommended.

For determining the audibility of the target staff, there are two options: either simply reading the audibility percentage value, which gives a rough estimate of the audibility, or taking all values from masking, centroid, and variation coefficients into account. The latter option requires experience in how different timbres act in orchestration and interpreting the values accordingly. For example, a bright glockenspiel sound has a very distinctive timbre and would probably be audible even if the orchestration masking curve covered the most of glockenspiel's partials.

Part III TESTING

1 WHY AND WHAT IS TESTED?

In this project, I research ways to resolve audibility problems of target instruments in orchestration as well as how to deal with the effects these problems have on my artistic decisions in composing. In the testing part of this report, I discuss my work in composing the opera *All the Truths We Cannot See*, which I composed while gaining new information about audibility and orchestration in general. I tested the usability of the Score-Tool App using the program as I intended it to be used: at the composer's working desk in the phase before rehearsals began. The main artistic output of the testing is the notated score and the performance of the opera. The discussion in this report includes examples of passages from the opera in which the Score-Tool App provided the most help. It also includes my own thoughts about how this project has affected my own composition practice. Besides the opera, I tested the Score-Tool App in my concerto for natural horn and wind orchestra. This testing gave me valuable feedback about how the Score-Tool analysis results correlate with live performances.

Another part of the testing is reporting how the Score-Tool App has been used by my peers, my fellow composers. The peer testing was done under my guidance, which enabled me to interpret the results of the Score-Tool App for other composers. The tested music includes works that had not yet been performed and works that had already been performed and showed some audibility problems in the performance.

The testing has given me information about how the masking algorithm is suitable for orchestration analysis, as well as about the experience integrating the Score-Tool App into the practice of composition. I have adjusted and developed the Score-Tool App code according to these testing results.

2 ARTISTIC TESTING: AN OPERA

Prior to starting my doctoral project, I had composed over 40 orchestral works, among which were four operas and two large-scale works for choir, soloists, and orchestra. In other words, I considered myself an experienced orchestral composer. For me, the most interesting aspect of orchestral timbre throughout my compositional career has been the impressiveness of the multi-octave chord. A chord composed with the careful registral and harmonic placement of individual timbres can create a sound effect comparable to nothing else – not a single acoustic nor an electronic sound. In all my orchestral compositions, I have sought the optimal setting to achieve the most resonant, most effective, and most full-bodied timbres called for.

A resonant, full-bodied timbre, as beautiful as it sounds, has a tendency to occupy the critical bands of the hearing system, overwhelming every hair cell with information so that nothing else comes through. This is sometimes a desirable effect, since it creates an immersive experience, drawing the listener irrevocably into the sound world that surrounds the senses. There are also cases when I want to draw attention to melody, the horizontal dimension of sound, and not only to harmony, the vertical dimension. A melody line, be it a vocal or instrumental soloist or just an important individual line in the orchestral part, is in danger of drowning in the timbre-flood where orchestral colors flourish.

Balancing between timbral brilliance and audibility has, for me, been a task of trial and error. In many of my early orchestral compositions, I could not hear the lines I wanted in the first rehearsals, simply because of a too-rich orchestration. Luckily, in some of the rehearsals, I had time to test different instrument and register combinations and began to build a tool bag of useful orchestration methods that suit my needs. After more than one hundred orchestral rehearsals of my own works, the tools I've collected are still not sufficient to orchestrate faultlessly so that I will get both the impressiveness I want and the audibility of the soloist.

Soloistic audibility is easy to achieve; just leave out as many instruments as possible from the orchestration or move them down in register until you hear the timbre you want. The other solution is to leave out the orchestration entirely when the soloist plays or back up the soloist with faint pizzicato in the double bass. This leads to a centuries-old layout for a solo concerto: the one versus the many, the struggle between David and Goliath, the musical drama where a soloist can shine as a hero before their colleagues. A symphony orchestra offers so much more than that. A carefully constructed orchestration lets the desired instrument's timbre shine along with the orchestration itself, delivering the best of both worlds, and creating the sense of filling up the soloist's sound so that it sounds bigger than it is. This is a task nearly impossible to achieve through trial and error because of the vast possible combinations and the different nature of every desired solo timbre.

In composing an opera, taking account of the individual singer's timbres is even more crucial than when composing for instruments, because of the variance of the sound material between singers. That variance can also be heard in the basic repertoire, where one singer can be almost inaudible in the same hall and with the same orchestra in which another singer comes through nicely. A ruthless audience can blame this on the singer for lacking the necessary strength of voice and for perhaps being overambitious in taking on certain repertoire. Every singer has their own vocal beauty, which can be respectfully underlined with an orchestration that does not call for superhuman abilities to achieve a successful performance. The singing voice can even be enhanced with orchestration that leaves the key frequency areas free for the singer to fill while still providing the extra body of timbre. In my experience, the extra body also makes the singer feel more secure than does singing with instrumentation that is too careful.

As fascinating and inspiring as orchestral timbre is, it is not the first parameter to consider in composing. Timbre is the flavor of musical tones, and the tones create the essence of a composition unless the aim is to focus entirely on timbre. Mine is not. In my aesthetics, music is constructed of closed forms of different dimensions; a piece is constructed of movements, movements are constructed of sections, sections of phrase groups, phrase groups of phrases, phrases of motives, and finally motives of notes. Each of these closed forms has its own dramaturgy, and together they form an entity in the same way as in literature alphabets, words, sentences, paragraphs, and chapters form a novel.

In starting a new composition, I usually have an abstraction of the final version of the sounding piece in my mind. The abstraction collides with reality as soon as I begin to write the actual music, that is, the notes with their properties, such as dynamics, articulation, playing techniques, and instrumental sound. The abstraction that appeared as a ready-sounding piece in my head was nothing but an unstructured collection of timbral and thematic ideas. The wholeness of the piece is impossible to imagine when there are no small-scale elements ready. The compositional process must be cut to easy-to-solve chunks of tasks, and solving each chunk helps to better understand the final goal.

2.1 PROGRAMMING AND COMPOSITION WORK

My doctoral project involved a great deal of computer programming, and the resulting Score-Tool App is the biggest single programming project I have done. During the process, I noticed how similar the programming work is to musical composition, especially in the process of designing a full-feature app. The end result is a working app with a user interface, but in the beginning, I had no idea either of the structure of the code or the steps I needed to take to achieve my goal. The first thing to do at the beginning of computer programming is to cut the over-sized task into smaller ones. Completing the small tasks takes you one step closer every time to understanding the overall structure.

In coding and composing, the end-result seldom matches the initial intentions. Solving small-scale tasks gives rise to new ideas, and perhaps creates problems, which affect the large-scale structure. In composition, for example, if the idea is that a large-scale structure imitates a small-scale form, say, a motive, in terms of harmonic rhythm, then changing the motive affects the large-scale structure. In coding, for example, if the idea is to visualize data as a line graph, but the function or method ends up returning a 3- or 4-dimensional matrix, the visualization needs to be changed. In addition, in composing I have a preference for a work to be coherent. Thus, even one small-scale structure that differs too much from the whole may require revising several musical layers. Both the composition and the coding practices are therefore an ongoing discourse between details and the whole, where both layers are under constant development until the premiere, and often even beyond that.

In doing my doctoral project, I switched between coding and composing during my workdays. I realized that the similarity in thinking in both practices created mutual help. I intuitively thought of the developed app program as an abstraction of a musical composition. In my mind, the program has a dramatic arc, i.e., the instructions start in silence (an empty data structure), develop through functions and methods, and end when the result is given to the user, after which the user perhaps starts a new cycle. A program also has themes and motives, as in a composition; small scale units solve problems, creating a sense of temporary relief. When I code, I think of the program’s classifications and methods as thematic entities; they have properties that can be changed only up to a certain point, after which the classification, or thematic entity, has drifted too far from its original purpose.

When I orchestrate music, the work reminds me of branching or parallel processing in coding. In orchestration, you have to deal with multiple simultaneous actions, which are dependent on each other. For example, it is important to consider the ability of instrumental groups that are normally seated spatially far apart to play complex rhythms together. The co-working of horns and double basses is not as smooth as that between oboes and bassoons owing to the seating distance between brass instruments. Similarly, in coding parallel processes, you have to keep in mind that one process may have to wait for data from another until it can continue. In orchestration, simultaneity is not always achieved by writing pitches at the same point in a measure. Likewise in coding, simultaneity is an illusion created by receiving consecutive data very fast. Both in orchestration and coding, efficiency is the key to the best performance. For an orchestra, it is not advisable to write long, loud tremolos for strings, because of the strain on the players’ hands, or long phrases in the upper register of the brass, because of the strain on the human breath. An equivalent effect might be possible with a lower “cost” to the musician. Similarly, in coding, large numbers of multiplications and divisions might cause delay in data processing. These can perhaps be replaced by bitwise operations, bringing the task “closer” to the processor.

The work of programming required me to learn new skills because of the fast-evolving nature of information technology. When I studied algorithms and database manipulation, I often thought of what the compositional equivalent of a data type or an operation would be. For example, if a certain kind of musical motive is equivalent to a list-structure, how could I make it more like an indexed list or a stack or a hash-dictionary? Is the programming API analogous to the compositional coherence? Is the difference between pointer and reference equivalent to the difference between allusion and quotation in music? Why do composers not leave comments in their scores in order to pass on information about the structure of their work, for example, by pointing out that this harmony is a flipped version of another, or that this motive repeats at slower rate than another in the movement?

2.2 THE COMPUTER AND CREATIVITY

If there are so many similarities between composing and programming, then one might ask, why don't I code a computer program that writes the music for me? The question might be generalized as a meta-question: why doesn't someone write a computer program that codes a program according to a simple wish, such as "Please code a program that composes a symphony for me." The answer is that, in order to do a task, a computer needs well-structured instructions about what the user wants. Computers are lousy when it comes to creativity and aesthetics. When a large-scale problem, like "writing a symphony" is divided into smaller tasks, it comes down to instructing the computer to decide what notes with what parameters are needed at a given point in time. To give these instructions to the computer, a programming language is the best way, but if the goal is a symphony, then the task is better performed by a human.

If a composer wants to give well-structured instructions for creating a symphony so that the result will sound the way composer intended, then the best way to give that data is to put it into a musical score. In that sense, a musical score is a computer program, and running this program is performing the score. If we write a computer program that writes a score, it would be analogous to writing a C-program that writes a java-program, for example. Extremely rare and inefficient. The benefit of coding along with composing is that it widens the horizon for how to handle musical data. After concentrating mainly on music for over a decade, this interdisciplinary project refreshed my attitude towards my usual routines in composition and has given me new perspectives on the standard orchestral repertoire.

The Score-Tool App does not touch the musical score in anyway. Rather, the purpose of the program is to act as "middleware" between the composer and the performance. Its main purpose is to help the composer in the orchestration of the piece. The program does not give direct answers for how to orchestrate or what to write, but it does give a first stage of feedback about the functionality of the orchestration. The program collects the data from a user's input, which is either a single orchestration chord or a full score, and provides a visualization of the data by putting the orchestration through mathematical models. The output is various visualizations, graphs, and numbers, whose interpretation is far from trivial, but is discussed both in the theoretical and in the artistic sections of this report. The theoretical part provides the tools to read the results, such as whether it is likely that the orchestration functions as intended or whether or not the target instrument is audible. The artistic part gives examples of how the results can serve as inspiration for the composer to make artistic choices which, without the program, would have gone undiscovered. The data visualizations act as an inspiration for a composer to experiment and play with orchestration colors and, instead of auditory feedback, get visual feedback. Auditory feedback can be deceiving, because the creation of an orchestra hall – like an auditory environment – is a hard task to perform at a working desk. In my view, visual feedback, created by algorithms with sufficient data to mimic the actual hearing experience, is the most objective way to present the orchestration. The graphs and values also give a composer something concrete to talk and think about in the field of musical timbre. As I stated in the background chapter, there is no generalized vocabulary for describing timbre. The concept of sound color is perhaps a misleading term, because the color names are not even used to describe sound color. The Score-Tool App timbre values and visualizations are efforts to make different timbres comparable. The App at least makes it theoretically possible to create orchestration that is “twice as homogeneous” or “half as distant” to the reference. As I mentioned earlier, these are meant to be inspirations for artistic choices, not scientific analyses of sound.

2.3 THE OPERA

All the Truths We Cannot See is an opera I composed on the side, so to speak, while coding and involved in acoustical learning in this doctoral project. The libretto was written by Glenda Dawn Goss. My knowledge of orchestration has grown while composing the opera, and the solutions I made in the early stages have been revised many times by now. There are still some hints in the opera score from the early stages, when I was orchestrating in my “old” style, without diving deep into the physical and acoustical properties of instrument sounds. An opera takes time to compose, and I could not wait to have my technical instrumentation apparatus ready before starting the orchestration. Reminiscences of early-stage orchestration appear, for example, as an optimistic choice of instrument groups in sections where the general mood calls for low-register singing. In these kinds of sections, it was sometimes difficult to change the entire orchestration, because the orchestration also affects the surrounding passages. Because the Score-Tool program alerts the user to audibility issues in some passages where the singer’s part is in a low register, I used these passages as test cases to see if the information was accurate. This report thus contains a comparison between the results of the program and my subjective experience in the live performances.

The impressiveness of timbre comes from the sensation that the listener is surrounded by sound, and the sound has the supporting body that makes the timbre convincing. The opposite of impressiveness is fragility and powerlessness, which are not to be taken as negative properties, because continuously powerful music creates fatigue. In my early orchestral compositions, I pursued the subject by arranging my orchestration to mimic the overtone series of a given harmonic tone. This has proven to be a good approach for achieving resonant timbre and is also recommended in orchestration handbooks.[246] A caveat to this approach is that it dictates the harmony, which was a minor question in the nineteenth century, but it makes an aesthetic statement in the twenty-first century.

2.4 SCORE-TOOL AND THE OPERA

Doing the work of composition while acquiring information about the perceptual aspects of music made me question my usual orchestration methods. The usual methods included creating a “registral tunnel” for the target instrument or singer so that the registers of orchestration and target do not collide. Browsing the analysis data of orchestral instrument sounds revealed that there are multiple cases where a low-register instrument creates a strong masking on a critical band two octaves above the root tone. Also, some of my target instruments have such a strong spectrum that leaving the frequency area of a notated tone free has little effect on the audibility of the instrument or vocal sound. In both cases, the orchestration problem could not be solved intuitively, but required analysis to reveal the true masking pattern in question.

My intention was to make a spectral analysis of the voices of each singer involved in the performance, a planned cast of international dimensions. But in the fall of 2020, when this plan was to have been put in operation, the COVID-19 pandemic ruined that idea because the whole opera production was postponed. Some of the singers moved to their home countries, and the rest understandably did not wish to meet in person. Instead, I used analyses of generic samples

---

[246] Adler 2016, p. 143.

of opera singing voices, mainly light/lyrical, which are not the exactly the data I had in mind, but these were better data than nothing at all. The generic voices are recorded using the vowel *a* at different dynamic levels. The voices have the basic opera singer's formant, which shows faintly in the graphs. The dramatic voices, which few singers in my opera project had, include presumably a stronger singer's formant, which is taken into account in my orchestration.

To ensure the audibility of each singer, a secure approach would be to prefer low dynamics and low registers in the orchestration. The two focal points of my opera project were the audibility of the singers and the development of my orchestration skills. Therefore, my aim was to have rich orchestration along with audible targets. Referring to Sundberg's vocal formant study,[247] a singer with a strong *squillo* (a ring or ping in the voice, see *vocal formant* in the Glossary) would automatically be audible in a texture with rich orchestration. In practice, the spectrum of operatic music is not as unambiguous as Sundberg suggests. Usually, the singer's voice does not peak as strongly at 3 kHz as claimed in Sundberg's study. Besides the *squillo*, there are other properties of the voice that need to be audible for the orchestration to do justice to the voice. In addition, the presence of *squillo* is questionable in the lowest and highest registers of the voice. Furthermore, the determination of audibility in music is more than just getting a hint of a sound that it is there. In opera, one might suggest ensuring the understandability of the sung text is also key, but that is beyond the scope of my project.

The idea of favoring rich orchestration in my compositions affected the work considerably. It also raised the question of listener's fatigue.[248] A constantly sounding rich timbre can be tiring on the ears, so I experimented with ways to have a rich orchestration with enough variety in timbre to avoid listener's fatigue. One thing I noticed was that even though nearly every instrument is otherwise in use, omitting the double bass from the orchestration for a while "refreshes" the timbre, so that when the bass re-enters, the timbre sounds rich again.

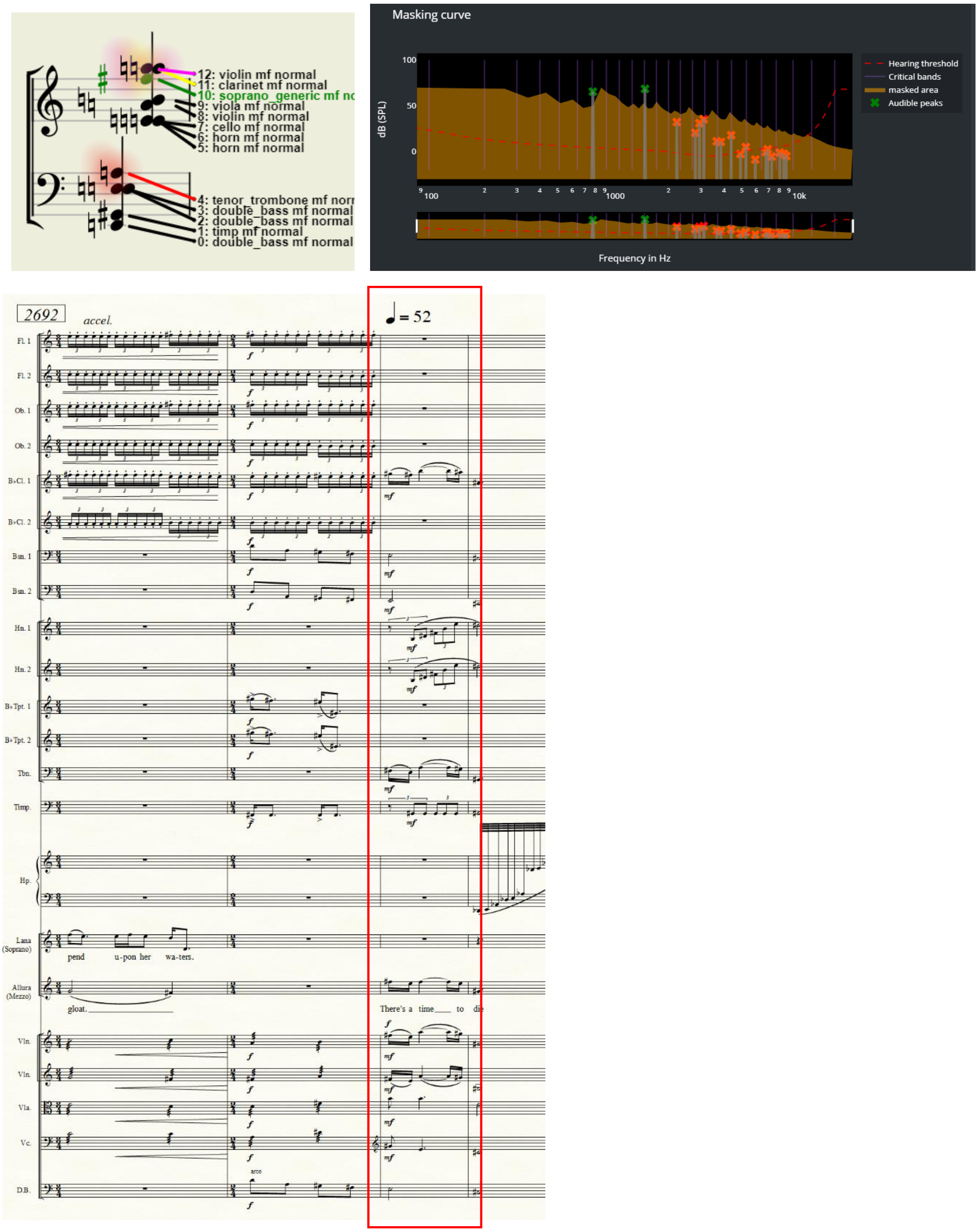
A good example of a section where rich orchestration is especially needed in the *All the Truths* opera is the duet by two female lead roles starting at m. 2627. During the scene, the orchestration thickens gradually, and at the same time the orchestra becomes more and more prominent. An analysis slice beginning in m. 2692 shows that the masking curve (which is also a representation of the spectrum, but with an applied spreading function) shows a triangular shape, with the apex at around 800 Hz. Based on Terhardt's idea that the area of spectral dominance area is located near at 700 Hz, this can be interpreted as rich orchestration.[249]

---

[247] Sundberg 1977.

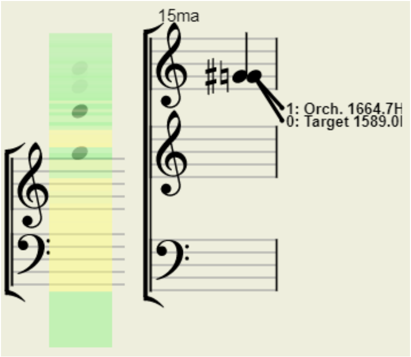
[248] Listener's fatigue happens when a pleasant sound begins to feel tiring, analogous to eating too much candy.

[249] See Part I, section 5.10, "Spectral dominant region."



Example 70. Uljas Pulkkis, *All the Truths We Cannot See*, mm. 2692-2695. In this graph the soprano is set as the target, and the program suggests that nearly all the singer's overtones would be under the masking curve. The true form of a singer's formant is seen here as an extra bump in overtone energy at the 3 kHz area, which in this case would not be enough to rise over the general masking threshold of the orchestration. Here, the more important aspect seems to be the second overtone, which is indicated to be about 20 dB louder than needed for audibility. Also, the surroundings of that particular overtone show that the masking curve is around -20 dB below the excitation point.

The masking curve projected on the musical staff in Example 71 shows that the critical bands appear unoccupied in the regions where the singer’s overtones are located. The spectral centroid data of both target and orchestration seem to collide around 1.6 kHz, indicating that both timbres are equally bright. This will somewhat diminish the audibility of the target, but not so much that it would render the +20 dB peak of the second overtone inaudible.



Example 71. The masking curve and target overtones in m. 2696 plotted on a staff along with the spectral centroid.

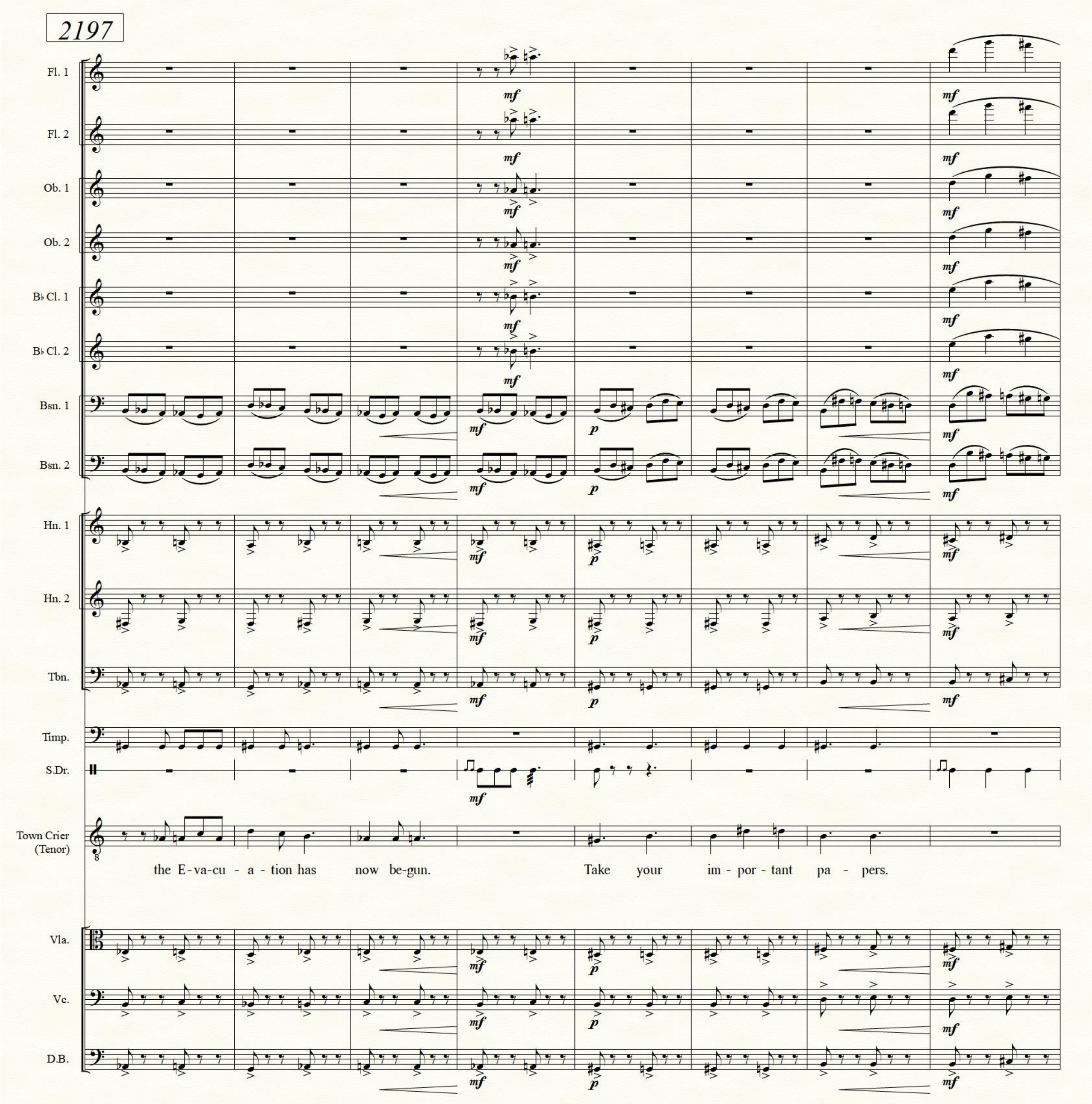
Taking the analysis data in m. 2692 (the red square in Example 70) as an example of the orchestration of the Act II duet in the opera, the orchestration seems to fulfill its purpose, both as being full-bodied timbre and avoiding masking the soloist completely. However, making even this type of coarse analysis during the actual work of composition might be cumbersome, and therefore the analysis data are intended as pre-evaluation information for the composer, who can make adjustments before the rehearsals.

The Score-Tool App allows us to consider orchestral textures as a combination of target and orchestration, since one of the main parameters of its program is the audibility of the target. In an opera or a concerto, the target is often self-explanatory, namely, the soloist, but in many cases the composer’s intention might be to create an orchestral sound entity without any pre-dominant tones. In composing this opera, I thought of some of the passages with singers in this way, by regarding the voice as a sound participating in the creation of rich orchestration. Especially male voices singing long tones can be thought of as pedals that hold the timbre together. The first overtones of the pitches in a typical tenor range reside conveniently just in the sweet spot of the rich orchestration, and thus can be used to create the essence of the sound while orchestral instruments fill in the surroundings of the formant in the spectrum. This technique also adds body to the singer’s voice if the formant register is kept free of frequencies other than the singer’s.

Having the data at hand during composition has made me rethink my usual routines in orchestrating music. In m. 2692, analyzed above, I began to think of the importance of having full-bodied timbre, and thus changing the horn texture from static to melodic in order to reach the important frequency areas for a rich orchestration. The notated root-tones of the horn in that register contain very few overtones, making them ideal instruments to create the spectrum peak at a desirable frequency. In this way, the use of the program affected the artistic choice I made even in this minor detail.

2.5 HOW THE SCORE-TOOL APP AFFECTS MY ARTISTIC DECISIONS

As I wrote earlier, in my previous operas I orchestrated mostly by trying to avoid registral clashes with the singers. This can be hard, especially when I want to use a rich orchestration. Many of the instruments have their best register in the same area as the singing voice, so avoiding that register could result in awkward parts for orchestral musicians. Therefore, I used to give instruments in the singing register low dynamics. As also mentioned, the masking issues in my previous compositions have been mostly caused by instruments playing in a register different from the target. That means I tried to avoid masking my target, but I did not have enough knowledge of all the factors that cause inaudibility.

Now, in composing the opera *All the Truths We Cannot See,* I noticed that I wrote for the orchestra more carefully than before. I marked the dynamics low, even for bass instruments because the masking phenomenon spreads upwards in frequency space. I was also extra careful about notating anything in the same frequency space occupied by the singer. Instead, I used a singer’s small pauses between phrases to notate for instruments that would have caused audibility issues for the voice. This technique creates a feeling of uninterrupted melody because a singer’s phrase continues in the orchestra. This can be seen in Example 72.

Example 72. In my opera *All the Truths We Cannot See*, the use of Score-Tool App had an impact on my artistic approach. I became more careful with dynamics, writing them even for bass instruments while a vocal soloist was singing. I also filled in a singer's pauses with the same frequency space the singer was using so that I could create the feeling of uninterrupted melody.

In this way, the Score-Tool App affected my artistic style and caused me to write with a technique I had not used before. In rehearsals, I also noticed that the singers were able to use a wider dynamic range than in my previous works. This was because the frequency space of their voices was free, so they did not have to compete with any other sound in volume.

I noticed as well that using the Score-Tool App made me concentrate more on the musical texture of the target instrument than I had done in previous works. In my pre-Score-Tool orchestral works, I thought of the orchestral timbre as one entity, even when writing a solo concerto. In other words, especially with concertos and vocal music, I considered the soloist a part of the orchestration, even though I wanted the solo to stand out. Using the Score-Tool App changed my view and caused me to regard the soloist as an individual timbre, a timbre that the orchestration supports and perhaps even enhances.

Using Score-Tool also slowed down my composing. In orchestrating passages with a vocal line, I re-thought all my decisions. I also exported the draft of the score I was working on into the Score-Tool App to check its recommendations. This was time well spent because it reduced the time balancing the orchestration in rehearsals, which was one of my goals in starting the whole project.

2.6 SCORE-TOOL IN MY COMPOSITION WORKFLOW

On a large scale, the Score-Tool App has given me freedom in my orchestration, as it enables me to evaluate the score already in the creating phase. Because the nature of the program is not to solve orchestration problems, but to analyze the composed orchestral score, my workflow has adapted accordingly. I use the program after spending considerable time composing a work to determine if the properties of my orchestration are meeting my intentions. Rapid feedback on orchestration functionality gives me the opportunity to write thicker timbres, louder dynamics, and denser texture than usual, since their functions can be verified already with the program, not in the first rehearsals with their busy schedules. In this way, I have found that the Score-Tool App frees up creativity, because all the minor details do not have to be dealt with immediately.

The homogeneity parameter, which estimates the diversity of timbres in a composition, has been another eye-opening parameter. The mixture of different timbres is hard to imagine as a sounding chord, at least for me. This is not a problem when using known good combinations, such as horn and cello or clarinet and flute in unison. The situation becomes more complex when combinations of different registers and playing techniques are used, and the result might or might not work the way one intends. The homogeneity value estimates the likeness of instrument timbres in an orchestration chord, and since the instruments with matching timbres are proven to blend well with each other, high homogeneity predicts a usable combination. The possibility of checking the homogeneity of any instrumental combination encourages one to try combinations previously avoided.

Originally, my intention was to add the homogeneity value to the program as a red flag for situations in which the target sound was about to blend into the orchestration. When I was composing, I noticed that I began using the “blending check” more often in positive way, such as setting an orchestration instrument as the target and ensuring that my instrument combination would blend with it rather than trying to make the target audible. The blending of orchestral colors in general has seldom been discussed in composition or orchestration textbooks. To my knowledge, the Score-Tool App is the first program to give a unit (homogeneity) and a value for this important parameter of orchestration.

2.7 EXPERIENCES IN OPERA REHEARSALS AND PERFORMANCES

2.7.1 GENERAL REMARKS

Opera rehearsals mostly take place with a solo piano accompaniment, the orchestra coming only to the last few rehearsals. Thus, singers get a false sense of the carrying power of their voices against the orchestration. This is exactly what happened in rehearsals of *All the Truths We Cannot See*, especially with the character Allura. In several of the first run-throughs, I was pleased to hear Allura singing at her full power when she first appeared on stage. However, a few days later, the same singer softened her expression considerably and said she had decided to make the scene more fragile. I had to explain to the singer that, in the scene, I used *tutti* orchestra and a full brass section. Therefore, the dynamic marking *forte* for the solo singer was justified, because otherwise the voice would not be heard. The singer had made the decision to soften her voice based on the piano accompaniment and would probably have switched back to *forte* after hearing the orchestra. Hearing her sing in the first run-through with the orchestra, I noticed that indeed the full singing power is needed for soloist audibility in that section, and the singer agreed.

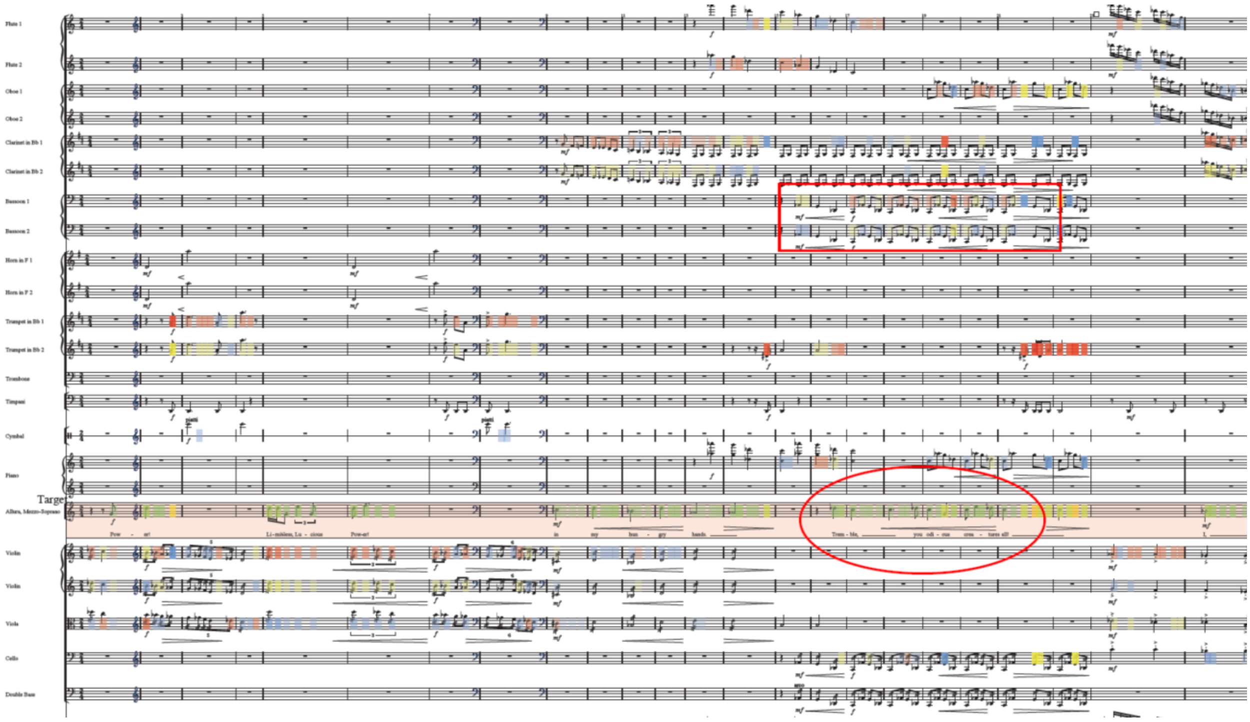
Another balance-related issue was the difference in voices between performers singing the same role. Most of the roles in the opera were double cast, but not with the same voice type. The biggest difference was between the leading soprano played in one cast by a dramatic singer and the same role in another cast by a lyrical singer. Surprisingly, there were no audibility issues, because even the lyrical singer had a strong voice. Other roles, however, would have required rebalancing the orchestration based on specific singers. That turned out to be problematic, because the orchestra would have had to write in several different dynamic markings in the same place in their parts and know which ones to use on which night. Therefore, I ended up balancing the orchestration dynamics based on the singers at the first performance. That led to some passages when the orchestra was playing too loud for the singers, although not to the extent that the voices could not be heard at all. Another solution would have been to balance the orchestration based on the lightest-sounding singers, but my intention was to keep the intensity of the music as high as possible, and thus keep the dynamics of the instruments high.

Balancing the orchestration appeared to be a trickier task than I thought. In the orchestral rehearsals, the players' interpretation of the dynamics was not what I wanted at first. For example, the dynamics *mp*, *mf*, and even *f* for trumpet sounded almost equally loud to my ears. Also, a string tremolo lowered the perceived dynamics noticeably compared to normal playing. For example, when *mf* changed into a tremolo on the same note, the perceived dynamic was *p*. The perceived dynamic of the lowest bass appeared to be problematic, because in many places only *forte* playing gave the desired effect to my ears. There was a little improvement when the double bass players moved to the center of the orchestra pit, near the back, which probably gave more prominent reflections of the sound than one would hear from their usual playing position on the side.

2.7.2 USING SCORE-TOOL IN REHEARSALS

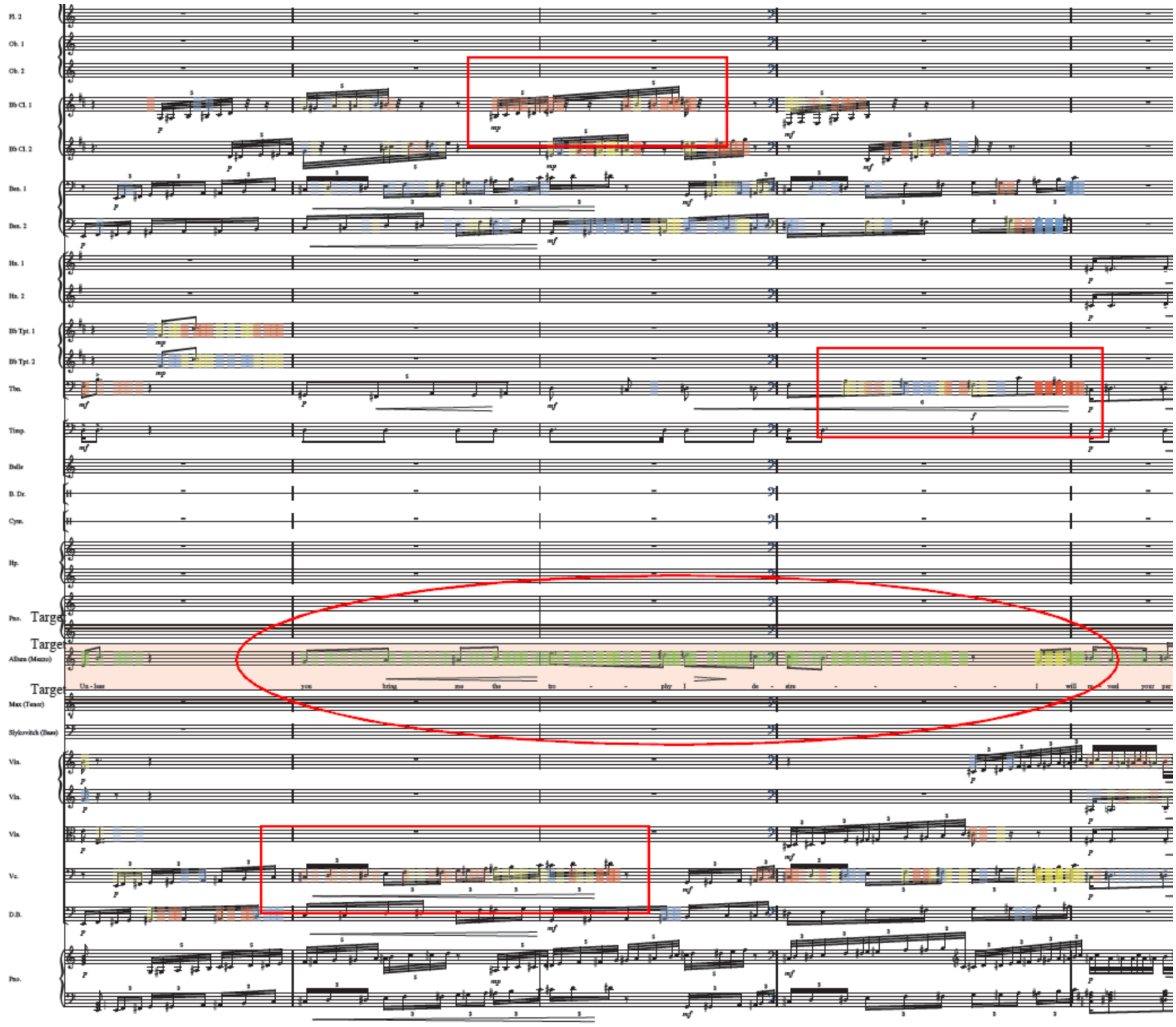
I ran the Score-Tool analysis of the score with parts for singers set as the target. In general, I checked that the audibility prediction percentage in the Score-Tool App was over 50 throughout the score. In some places, especially at culmination points, that percentage was occasionally lower than 50, when brass instruments were marked with *forte* dynamics. Because there was almost a week of rehearsals for the orchestra with singers, I had the chance to compare the Score-Tool data with the actual performance without hurry. Here are a few examples of how I used the Score-Tool App in those rehearsals.

In the character Allura's opening scene in Act I, there was a moment where the mezzo soprano's voice was suddenly inaudible for a short period of time. In this place, seen in Example 73, the Score-Tool App gave an audibility prediction of 60%, shown in the circled area in the example. In composing this passage, I had left it and thought that there would not be any issues. However, in the rehearsals, I immediately had a solution at hand because the Score-Tool App indicated that bassoons (shown in the square in the example) might be the problem. Lowering the dynamics of the bassoons solved the inaudibility issue.



Example 73. Excerpt from the *All the Truths We Cannot See*, Allura’s entrance, Act I. Bassoons caused audibility issues for the mezzo-soprano.

In Act II, there was another place with similar issues for the same singer, indicated in a red circle in Example 74. Here too the Score-Tool App did not alert me to any inaudibility, but it did indicate instruments with strong-sounding overtones in the same frequency region as the target sound. In this case, the Score-Tool information appeared to be valuable, because different instruments were causing trouble in different parts of the score. The three instruments masking the singer were the cello, the clarinet, and the trombone. The relevant places are shown in red squares in the example. Here the solution was to ask the soloist to sing a bit louder in the beginning, because the masking instrument was already playing *piano*. The clarinet dynamic level was kept at *p*, and the *crescendo* and the *forte* omitted from the trombone. With these adjustments, the masking effect was not as strong as before. Note that here timpani and bassoon kept their written dynamics.



Example 74. Excerpt from the *All the Truths We Cannot See*, Act II, with singers for the characters Allura and Max. The audibility of the mezzo-soprano is affected by the cello, clarinet, and trombone.

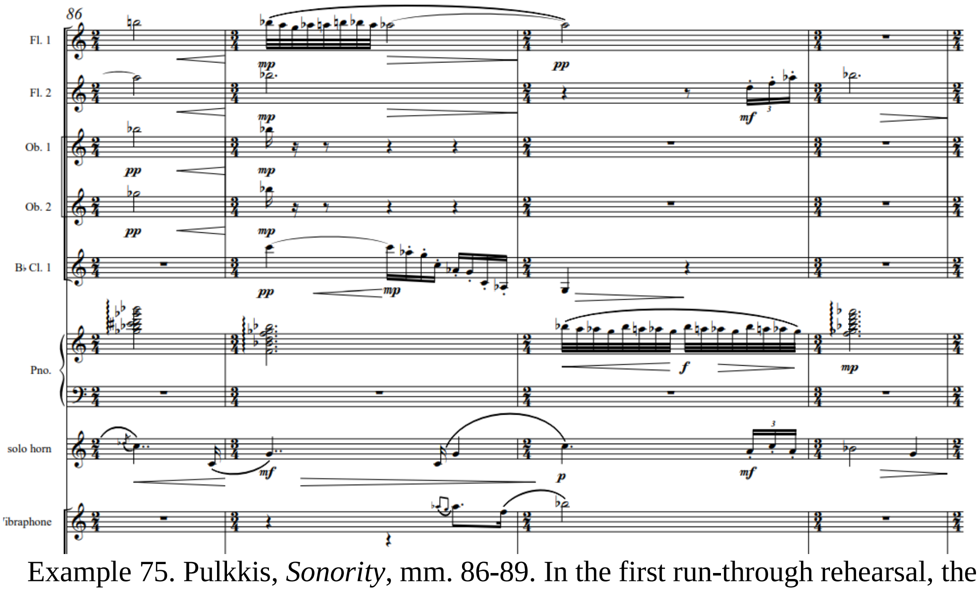
3 PEER TESTING

In developing the Score-Tool App, it came apparent to me that an app is like a new instrument; it requires much training before it gives desirable results. There are many things one must understand about psychoacoustics and timbre to be able to interpret the App's graphs and values and their role in the analysis results. It would be unreasonable to ask my fellow composers to devote their time to study the subject so that they could use the App in the same way I did in composing. That is why the peer testing included my own horn concerto and some short excerpts from other composers. Peer testing gave me many ideas about how to improve the App and its user interface in the future, but this is not within the purview of my doctoral project here, as my focus is not on user interface.

3.1 *SONORITY* – HORN CONCERTO, NOVEMBER 2, 2021

Sonority is my composition for solo horn and wind orchestra. The solo horn part is written for a natural horn, i.e., a horn without valves, with the exception of a passage in the middle of the piece where the soloist changes to a valve horn for a few minutes. I composed this piece by checking the soloist's audibility from time to time with the Score-Tool App. The work begins and ends with "secure" orchestration in which the soloist is clearly audible. In addition to those moments, I used the "maximum allowed" orchestration according to the Score-Tool App while keeping the solo horn sound just above the masking threshold.

I had the chance to hear a run-through rehearsal over a month before the actual performance. In this run-through I immediately reacted to the fact that dynamic markings are not intuitively interpreted as absolute values by many orchestra musicians. In Example 75, the woodwind players used a much higher dynamic level than was indicated in the score. I intended the *pianissimo* to be as soft as possible, and the *mezzo-piano* at a level well below *mezzo-forte*. In the rehearsals, the *pianissimo* was closer to what I call *mezzo-forte*, and the *crescendo* to m. 87 sounded almost like *forte*. In this first rehearsal, the soloist was not present, so I could not immediately hear how the rise in dynamics affected the soloist's audibility. However, I asked the orchestra to play very softly, especially in the places marked *pp*.



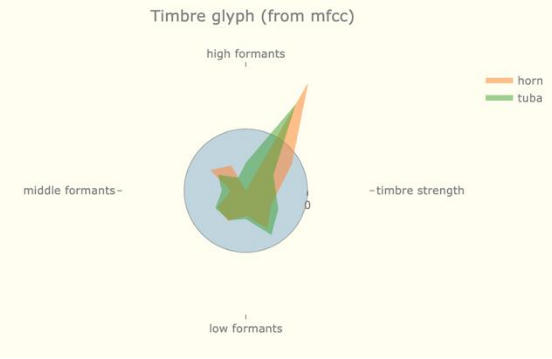
Example 75. Pulkkis, *Sonority*, mm. 86-89. In the first run-through rehearsal, the players interpreted the soft dynamic markings louder than I intended.

Before the first performance, we checked the solo part with the soloist and talked about the importance of following the dynamic markings. I noticed that the tones played as *stopped* (with the hand fully inserted into the bell) were softer than I had expected. I made the final changes to the orchestration dynamic markings to ensure the audibility of the stopped notes.

During the rehearsal with the whole orchestra and the soloist present, I finally heard how the orchestration functioned with the soloist. The rehearsal space was an auditorium with a low ceiling, which made it difficult to hear the real balance of the orchestra. In the auditorium, my listening position was by the conductor's side, about two meters away from the soloist. In the rehearsals, I reminded the conductor several times about following the given dynamics, especially when the marking was *p* or *pp*, but in general the horn was audible, although the wind orchestra's timbre was very similar to the soloist's from time to time, which made the soloist's sound blend into the orchestra.

The day of the concert was the first time *Sonority* was performed in a concert hall. To my surprise, the soloist was drowned by the orchestration in many passages, even in places that had been audible in the rehearsals. An excerpt of one of these passages can be seen in Example 77. Here, the dynamics of the woodwinds are *forte*, but there should still be room in the frequency space for the soloist. Sitting in the back row of the concert hall during the performance, I noticed that the dynamics of the soloist were partly masked by the orchestra. Apparently, in the rehearsals my listening position had been inside the critical distance of the auditorium, where the solo horn sounded louder than outside the critical distance.

In my view, there were at least two psychoacoustical features in the solo horn sound that also affected audibility. One has already been mentioned: the similarity of the horn timbre to the orchestration. In *Sonority*, I had called for two tubas and a baritone horn, whose timbres are close to that of the open horn. In addition, the saxophone timbre resembles the timbre of a stopped horn. The timbre similarity, which can be seen in Example 76, can make the soloist sound like part of the tuba section.

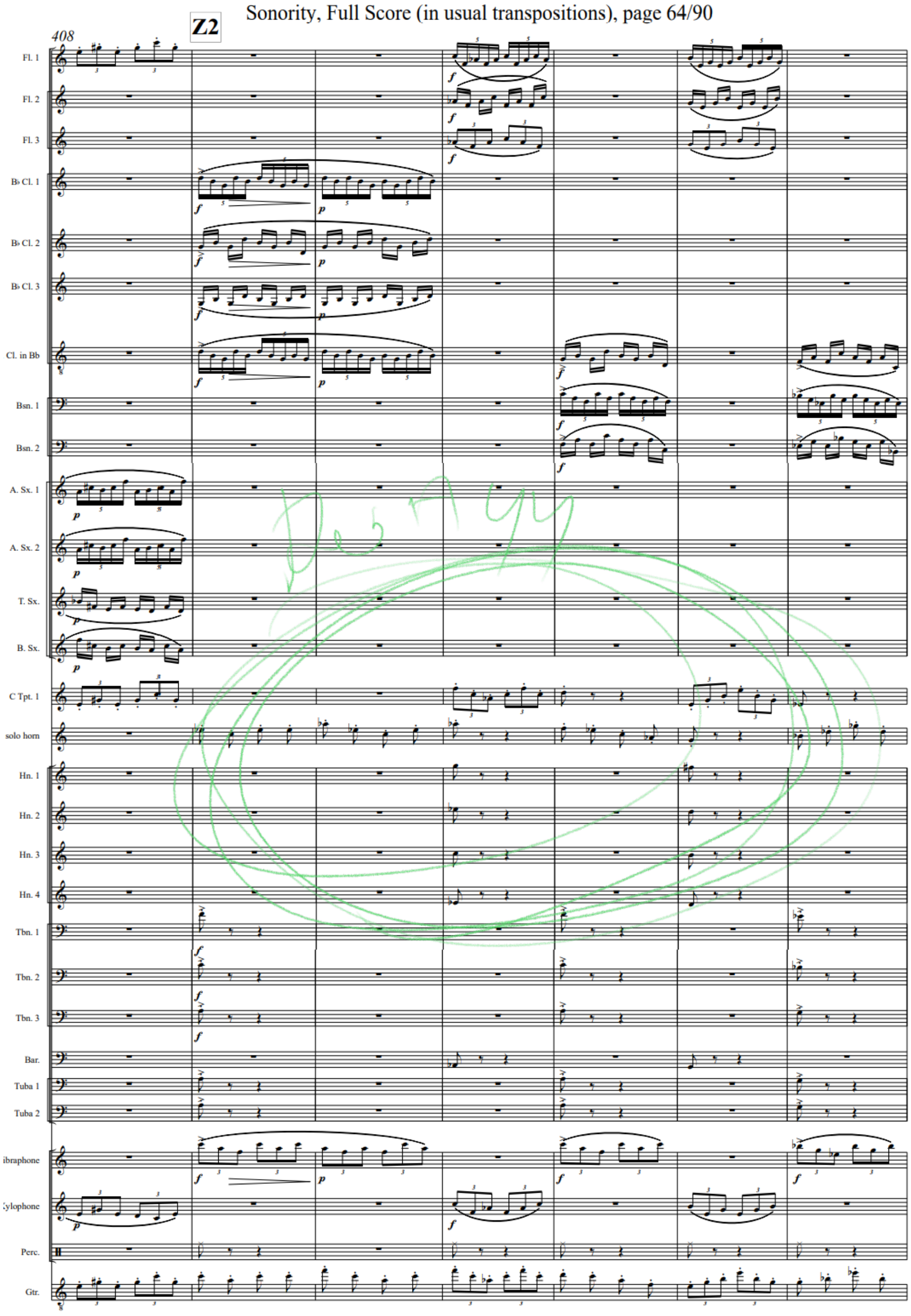


Example 76. The timbre graph,[250] taken from the Score-Tool App, shows the similarity between horn and tuba sounds. Both have strong timbral strength, and negative values for most of the formant areas, indicating that the sounds create a strong sense of a fundamental tone with no emphasis on any frequency areas above the fundamental frequency.

---

[250] The timbre graph is from the new version of the Score-Tool App. The graph is my own innovation to describe timbre: a circular presentation of the MFCC vector. Here, matching graphs equals a matching formant structure, i.e., a matching timbre.

The other psychoacoustical feature significant to audibility is the spectral centroid of the horn sound, especially the open horn. For example, the spectral centroid of the sounding tone *a3* (notated *e4* for the horn in F) is approximately 1.5 kHz. For comparison, the spectral centroid of a trombone playing the same tone is 3.5 kHz. As I described earlier in the Blending chapter, a low centroid is the single most significant feature of timbre that results in a timbre blending into other sounds. In *Sonority*, while the sound itself is perhaps not masked, the low centroid makes it difficult to distinguish the soloist from the mass.

With some adrenaline in the blood in a performance, orchestra players might also unintentionally play louder than in rehearsals. At the composition stage, I thought that when the Score-Tool App indicated that the soloist sound was more than 3 dB above the masking threshold of the orchestration, it would still be audible. This was apparently too tight a margin for a solo concerto: in a concerto, listeners expect the soloist's sound to be prominent, not just barely audible.

Example 77. Uljas Pulkkis, *Sonority*, mm. 408-414. The solo horn part was almost inaudible in the concert, although it was audible in the rehearsals. The Score-Tool App indicated that, performed with the given dynamics, the spectral peaks of the horn are 3-6 dB above the orchestra’s masking level in this passage.

Based on my experience with *Sonority*, I would say that just playing the low dynamics specified in the orchestration is not enough to ensure the target’s audibility. At the composition stage, while using the Score-Tool App, it might be wise to ensure that the target is also audible when the dynamic levels of the orchestration instruments are higher than indicated in the score. This would ensure that even in the excitement of performance, audibility is not in danger. This could be achieved, for example, by altering the octave registers of the orchestration or by choosing orchestral instruments with a low spectral centroid.

In the case of *Sonority*, I first realized the need to alter the orchestration only on the day of the concert, which was too late to make new parts for the whole orchestra. However, this experience led me to alter the audibility algorithm in the Score-Tool App. In the current version, the audibility decreases gradually when the target's spectral centroid is below 2 kHz. Also, a similar gradual decrease happens when the target's timbre, i.e., the MFCC vector, is close to the orchestration's MFCC vector.

Example 78. Uljas Pulkkis, *Sonority*, mm. 409-415, solo horn only. Re-analysis with the updated audibility algorithm of *Sonority*'s solo horn part shows that there are several warnings on the App about audibility. The green color indicates audibility of 50% or more, the yellow indicates only 25% or less audibility, and the red under 10% audibility.

Analyzing the score of *Sonority* with the updated algorithm of the Score-Tool App reveals that taking both the spectral centroid and the timbral likeness into account, the App indicates problems in audibility, as can be seen in Example 78. This indicates that the masking pattern alone is definitely not sufficient to determine the audibility of the target instrument in the orchestration, but the timbral features must also be taken into account.

In addition, the target's spectral peaks above the orchestration masking threshold do not automatically guarantee a target's audibility. Even the slightest dynamic variations in performance can alter either the orchestration's masking pattern or the amplitudes of a target's overtones. Thus, the Score-Tool's analysis is not as accurate as I hoped it would be. Therefore, it might be wise to test the target's audibility both with low target dynamics and with loud orchestration dynamics, especially if the target's dynamic marking is *forte* in the score.

3.2 Field test 2, September 28, 2021

3.2.1 Subject

A highly experienced composer with an already performed flute concerto that had audibility problems.

3.2.2 Motivation

To determine if the Score-Tool App would provide help in future performances.

3.2.3 Meeting

A private meeting with the composer, with the program running on my laptop.

3.2.4 General remarks

The use of the program was not clear to the composer. The composer had extracted a passage from the score using different software than mine. The score did not load into the Score-Tool App right away, and it was necessary for me to change the file type to xml. Thus, the composer could not use the program without assistance.

3.2.5 Checked excerpts

The composer selected one passage in which most of the audibility problems occurred. The problematic instrument was the solo flute, which was heavily masked by the orchestration during rehearsals of the piece. In the problem passage, the orchestration is quite thick, but the solo flute plays in a high register where the instrument can produce a piercing sound.

3.2.6 Passage

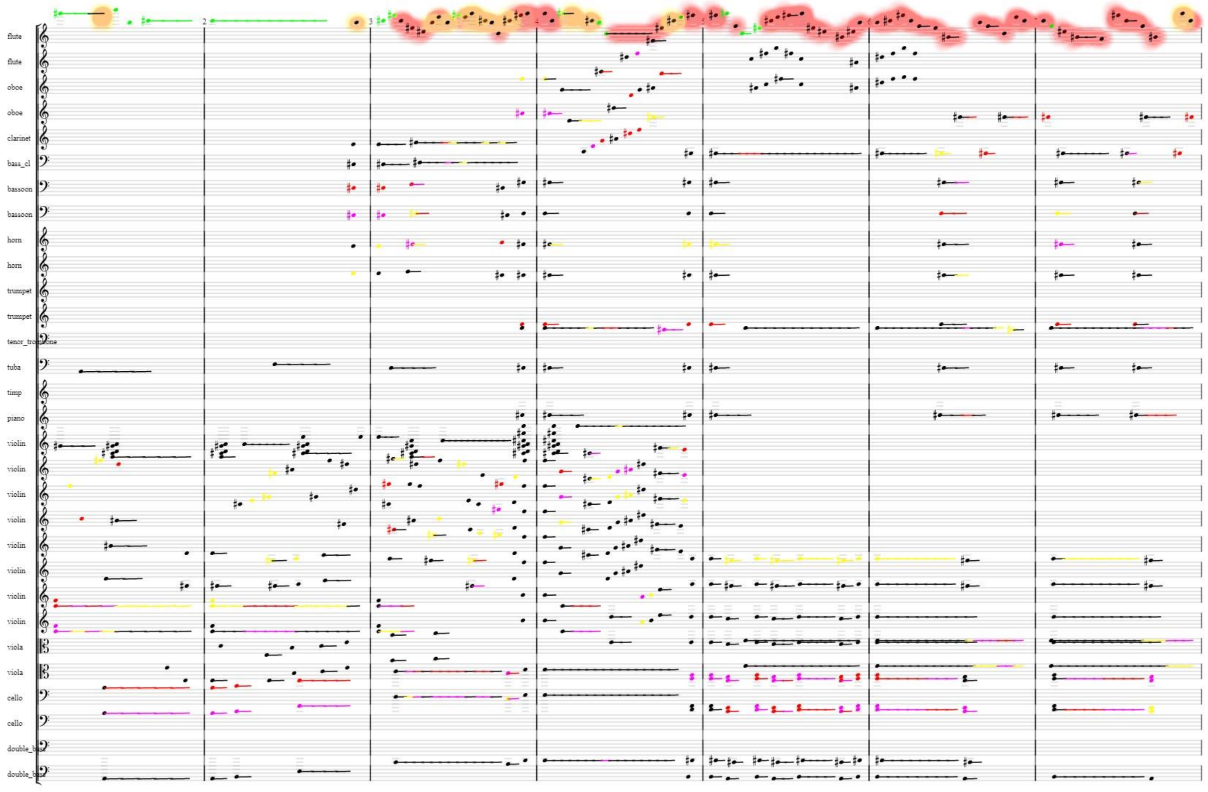
The passage is written for *tutti* orchestra in a quite conventional manner. The soloist plays rapid virtuosic figurations, while the orchestration is dominated by strings playing in the flute’s middle register. Brass, bassoon, and piano are playing short percussive accents with the exception of the trombone, which plays long tones with *glissandi*. The remaining woodwinds add short, rapid comments, but do not play constantly throughout the passage.

The composer explained that in the rehearsals, the conductor pointed out that he could not hear the flute. The composer said that he did not notice the problem because he already knew the piece and could imagine hearing the flute, although its sound was masked by the orchestra. The soloist also had trouble playing the example, because the musician felt uncomfortable playing a difficult passage knowing that the flute’s sound was being drowned out by the orchestra.

The conductor suggested that all the string parts be switched to solo strings, so the overall sounding mass would be smaller. The composer agreed to that, but the solution was not sufficient to resolve the problem. Furthermore, the conductor asked the strings to play softer than indicated, because the string parts contained *forte* dynamics. With these corrections, the solo flute became audible. The composer said that, because of a tight rehearsal schedule, there was not much time to spend on this particular passage. Otherwise, other solutions could have been tried.

The composer said that while composing this passage, the audibility problems with the flute were not apparent. He had thought that scoring the flute part as the highest instrument in the register would be sufficient to ensure audibility. Therefore, the audibility problems in the rehearsals came as a surprise.

In checking the passage with the Score-Tool App, we found that the program painted the flute part with a yellow-and-red background, indicating a warning for audibility, as seen in Example 79. This correlates with the experience the composer had in the rehearsals. A somewhat surprising analysis result is that in the latter half of the passage, only the cellos are marked as heavy maskers. This indicates that the conductor's decision to change all string section parts to solo parts might not have been the optimal solution to resolve the soloist's audibility issue.



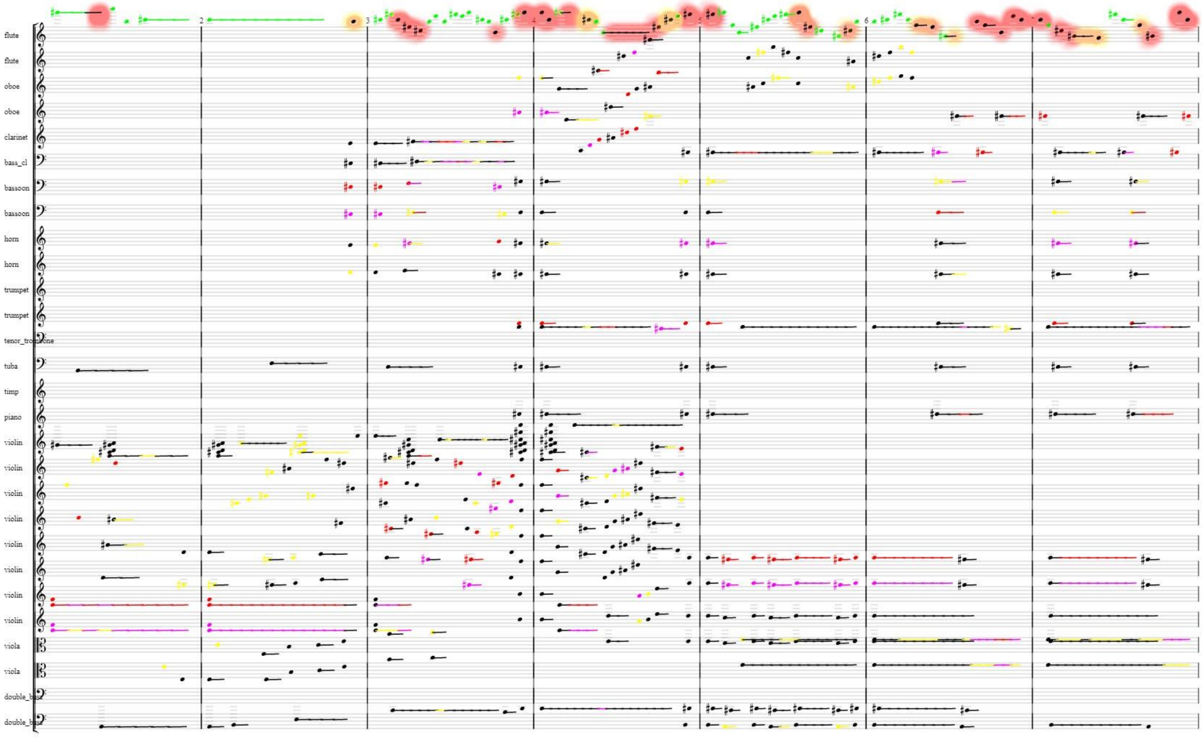
Example 79. Anonymous, *Flute concerto*, passage with soloist and orchestra. A masking graph of the score. The solo flute staff is at the top, and Score-Tool shows a red-and-yellow background with most of the soloist's notes, indicating that the solo flute is probably inaudible. In the latter half of the passage, cellos are identified as the strongest maskers.

The result was discussed with the composer, and after seeing this result, he recalled that in the rehearsals the conductor had also said something about the cellos, but the composer couldn't remember exactly what that was.

We continued the Score-Tool analysis by clicking the individual soloist's notes to reveal the orchestration chord at a given time. In Score-Tool, I tried to lower the dynamics on the cellos in several orchestration chords. That lowered the height of the masking curve in the graph, so that more partials of the solo flute were above the curve, indicating that they would be audible.

After this, we decided to make another calculation of the whole score by revising the dynamics of the cellos to *p*. That resulted in the solo flute part having a yellowish color instead of red, which meant that the audibility increased, but the issue was probably not entirely resolved.

We made one more calculation by excluding the cellos from the orchestration altogether. That resulted in the solo flute's color background turning to green in many places, indicating good audibility, as seen in Example 80.



Example 80. Anonymous, *Flute concerto*, passage with soloist and orchestra. Audibility test. Excluding the cellos from the calculations made the solo flute part audible in several places. The original situation with the cellos can be seen in Example 79.

There remained some red places in the flute part, where the heaviest masker was often indicated as the violin or viola. We did not go into details to make the whole flute passage show green, because the composer was already pleased with this result in which the solo flute part was often audible.

The composer said that in the next rehearsals the cellos could be asked to play the passage as soft as possible. It could then be checked to determine if that would be a sufficient solution to correct the audibility issues. After knowing the analysis results, the composer was a bit surprised that a low-range instrument could be guilty of masking, but after thinking about it admitted that this was probably the case. The composer said that the Score-Tool was a useful device for solving the audibility problem and would use the App in the future compositions.

3.3 FIELD TEST 1, JANUARY 30, 2020

This testing of the algorithms was done with the alpha version of the Score-Tool App, which I programmed with Matlab. The graphs look different from the current version of the App, but the underlying algorithms are the same.

SUBJECT

An experienced composer with a new work for chamber orchestra.

MOTIVATION

To check a few orchestration excerpts with possible masking problems.

MEETING

At the composer’s studio to which I brought my laptop. The program was running on my computer.

GENERAL REMARKS

The use of the program was not at all clear for this first timer. My presence was essential for the test to work. The score navigation according to the piano-roll view[251] appeared to be difficult, even for me with an unfamiliar score. The composer was familiar with the basic concepts of the sound spectrum, masking, and critical bands, but the meanings of the graphs in my program were not clear at first glance. When I explained aloud what one could see on the graphs and the significance of the numbers, the composer seemed to understand the parameters.

CHECKED EXCERPTS

The composer selected five discrete passages from the score, where the audibility of a specific instrument was important. Of these passages, two were more or less self-explanatory (such as an oboe solo with thin orchestration), but three were interesting test cases for my program. Of the remaining three passages, we decided together on a few excerpts to examine closely.

FIRST PASSAGE

The passage included a colorful and registrally widespread orchestration. There was a slight mismatch between the orchestration and the database. The composer used a *Bartók pizzicato*, which we replaced with ordinary *pizzicato*, and *flatterzunge* was replaced by *tremolo*. The composer wanted to know if the bassoon could be heard. In the passage, the bassoons play a tritone in a low register. That version of the program did not support multiple targets, so I decided that we set Bassoon 1 as the target and turn Bassoon 2 off. In the rehearsals, the other bassoonist was absent, so our choice was a lucky one. The score can be seen in Example 83.

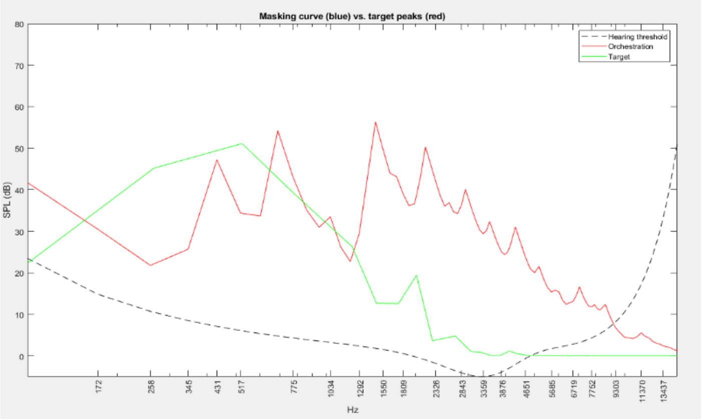
---

[251] piano-roll view is a common way to show a midi file. The view consists of a picture of a piano keyboard and lines indicating which key is playing and for how long, like in rolls in player pianos in the late 19th century.

The masking graph of the score (Example 82) showed some green areas for the bassoon sound, and clicking the mouse indicated good target audibility on a few of the 106 used bands. Note that the low bands are at the top and the high bands are at the bottom.

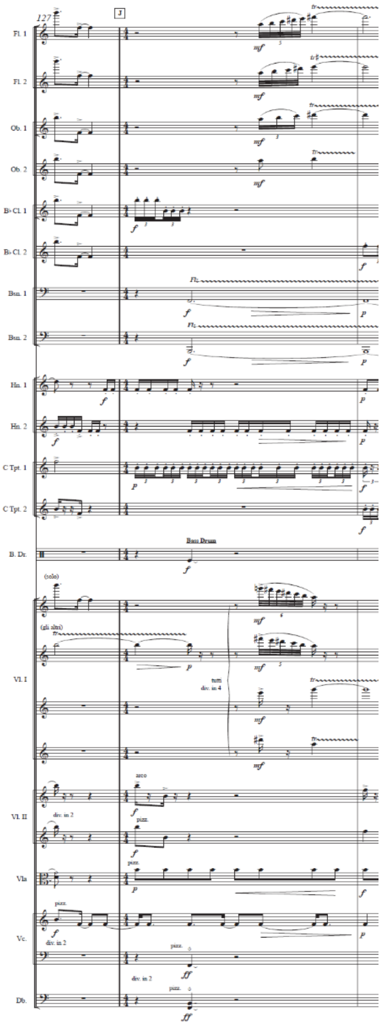
Critical band frequency: 517 Hz
Orchestration masking threshold: 16 dB
Target critical band excitation: 45 dB
Target to orchestration ratio: 29 dB
Good target audibility on critical band

Example 82. Anonymous 2, work for chamber orchestra, rehearsal letter J. The masking graph is equivalent to the 3D graph in the current version of the Score-Tool App. The masking graph of Bassoon 1 against the orchestration. A mouse click on band 517 Hz reveals that the sound can be heard on a lower critical band in our hearing system.



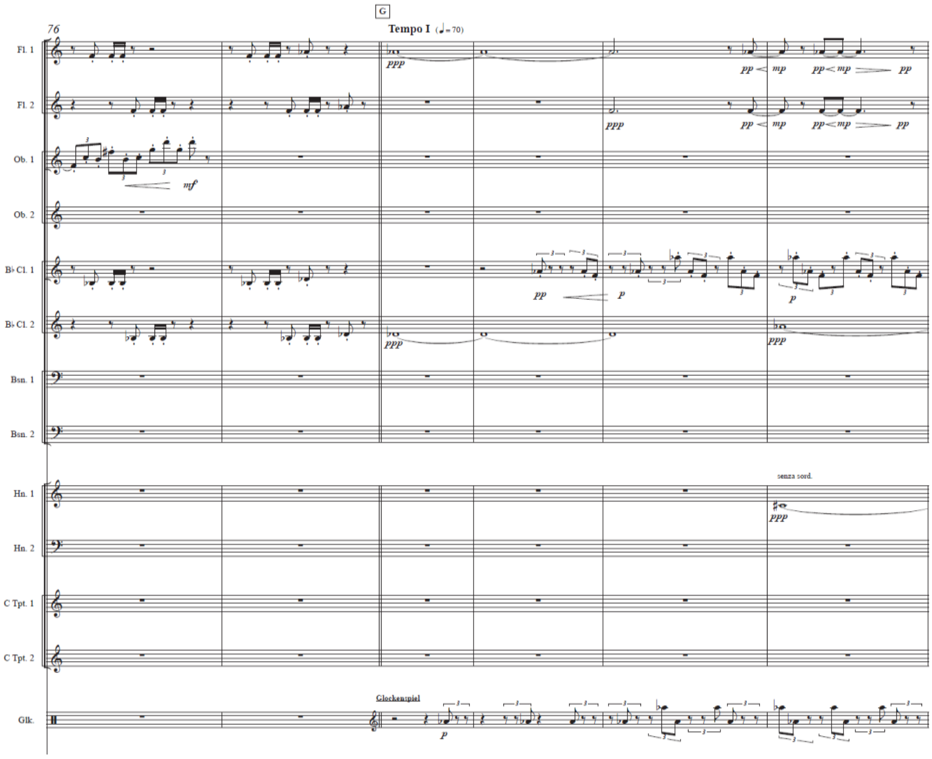
Example 81. The masking graph of the orchestration (red) against the loudest peaks of the bassoon (green). The dashed line shows a general hearing threshold.

The expert view (the same as the graph view in the current version of the Score-Tool App) in Example 81, in the same place as the previous mouse click in Example 82. Anonymous 2, work for chamber orchestra, rehearsal letter J. The masking graph is equivalent to the 3D graph in the current version of the Score-Tool App. The masking graph of Bassoon 1 against the orchestration. A mouse click on band 517 Hz reveals that the sound can be heard on a lower critical band in our hearing system. shows that some components of the bassoon sound are indeed over the masking curve, but some of the louder overtones are heavily masked by the orchestration. The orchestration variation coefficient showed the number 0.516772, indicating that the orchestration timbre would be on the homogeneous side. Based on the data, I assumed that the low sound could be heard, but that the bassoon would blend in.

After the rehearsals, the composer reported to me on the first passage as follows:

Example 83. Anonymous 2, work for chamber orchestra, rehearsal mark J. The target instrument is Bassoon 1 (7^{th} staff from the top). Bassoon 2 was turned off in the program.

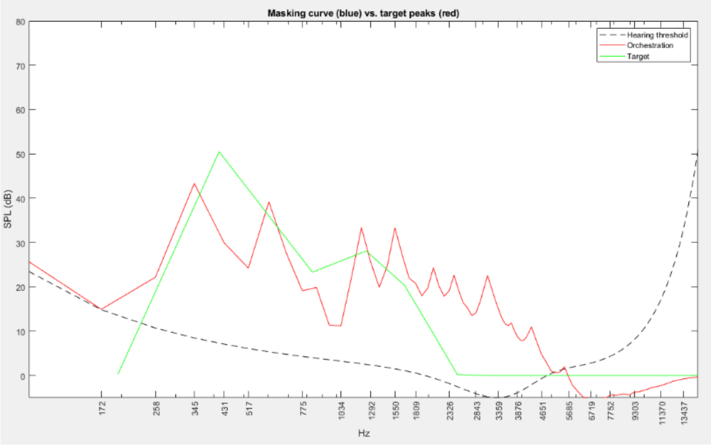
> “Matalien fagottien sulautuminen: Tässä haittasi toisen fagotin puuttuminen, joten homma meni hiukan “laimeaksi" jo tästäkin syystä. Matala fagotti kuului aika hiljaisena ja pyöreänä (ei fagottina), mutta kuului. Meni aika tavalla ennustuksen mukaisesti. Mainittu “laimeus" johtuu toki myös matalan tritonuksen puuttumisesta.”
>
> “The blending of the low bassoons: Here the absence of the other bassoon disturbed, so the thing went slightly ‘mild’ already for that reason. The low bassoon could be heard quiet and round (not like a bassoon), but it could be heard. Went more or less like expected. The ‘mildness’ I mentioned is caused, of course, by the absence of the tritone.” (Translation by the writer)

3.3.1 The second passage

In this passage, the composer wanted to determine whether the flutes were being masked by the orchestration. Here I found an unfortunate glitch in my program. Even though the dynamics in the score are marked at low levels, my program assumed that they were *mf*. Therefore, the general graph showed wrong results and was useless. We switched right away to Expert view and examined the middle of the last measure in Example 84.

Example 84. Anonymous 2, work for chamber orchestra, rehearsal letter G. The target instrument is Flute 1 (first staff at the top).

In the Expert view, I changed the dynamics of all the instruments to *piano*, as it is the lowest dynamic setting on the database. This does not represent the score excerpt very well, as the dynamic markings there vary from *ppp* to *mp*. This required some explanation on my side about the graphs in this situation; we are not investigating the actual orchestration, but a simulation. The spectral content of an instrument sound may not vary much when played *ppp* versus *p*, but the perceived loudness may vary by several decibels, and thus the graph here is an approximation.

The masking graph shows a strong peak for the notated pitch of the flute. The overtones are masked by the orchestration, although something from the overtones around 1kHz could be heard. The orchestration variation coefficient shows a quite high number, 1.76251, at this orchestration, indicating non-homogeneous color. The program estimated that the strongest masker would be the first violin. Based on the data, I assumed that the “fluteness” of the sound could be hard to perceive, but the notated tone could probably be heard.



Example 85. The masking graph from the middle of the last measure of the Example 84. The orchestration is on red and the target on green. The hearing threshold is the dashed line.

After the rehearsals, the composer reported to me on the second passage as follows:

> “Huilut väliäänissä: Aluksi huilujen komppi (paisutukset) erottuivat, mutta ne jäivät myöhemmin jousien ja oboen alle. Muistiinpanoissani lukee, että välillä viulut peittävät, joten kyllä. Jotain toki kuului, mutta säestyskomppia ei oikein erottanut jousien ja oboen alta. Ennustuksen lupaama “huilumaisuuden” katoaminen mielestäni toteutui: Neutraali pyöreä ääni, joka lähestyi hiljaista käyrätorvea.”
>
> ”Flutes in middle voices: At first the riff of the flutes (swells) could be heard, but they remained later under strings and oboe. My notes say that sometimes violins mask, so yes. Something could be heard, but accompaniment riff could not be clearly heard under strings and oboe. The suspected disappearance of ‘fluteness’ came true from my point of view: A neutral round tone, which came close to a quiet French horn.” (Translated by the writer)

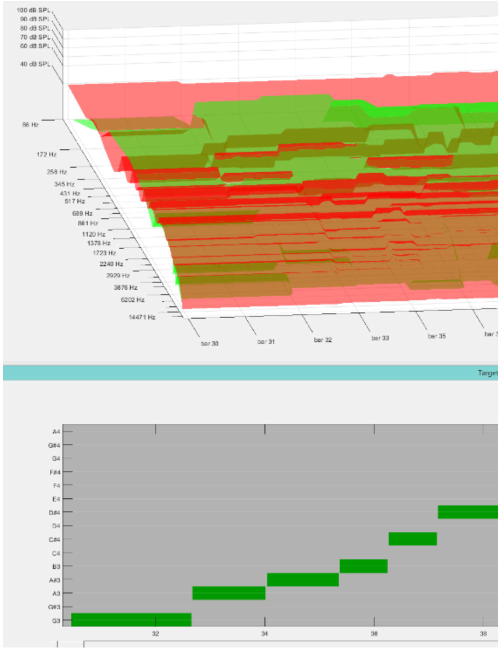
3.3.2 THE THIRD PASSAGE

The third passage was an orchestral *tutti*. The score can be seen in Example 88. The dynamic marking for all instruments is *ff*, and the composer wanted to know if the melodic line of the French horn could be heard.

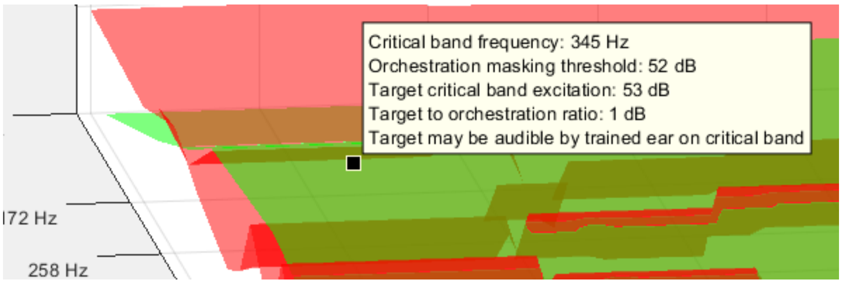
In the program, the dynamic marking *f* was automatically assigned to each instrument. This might give a slightly inaccurate result for the brass instruments, which emphasize the overtone content at high dynamic levels.

I noticed the problem immediately with the target instruments in unison. In future versions, there is a need for a “unison” switch, which raises the curves by 3 dB if there is a second instrument similar to the target playing in unison. In this situation, I was unsure how to correct the graph, so we used the results as if there was only one French horn in the orchestra.

In this passage we first noticed that the 3D graph of the passage showed green (Example 86), but the peaks appeared to be just a bit over the masking threshold of the orchestration. Clicking the mouse on the green area (Example 87) showed that the spectral peaks are just 1 dB over the masking threshold, which indicates that it is difficult to tell if the target sound is audible.



Example 86. Anonymous 2, work for chamber orchestra, rehearsal mark A. The 3D graph. The lines show the correspondence of the French horn piano-roll tones to the graph.

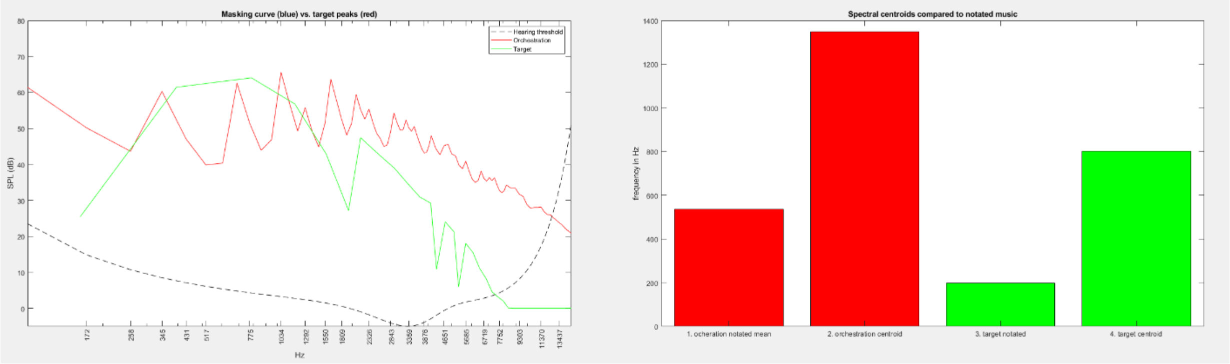


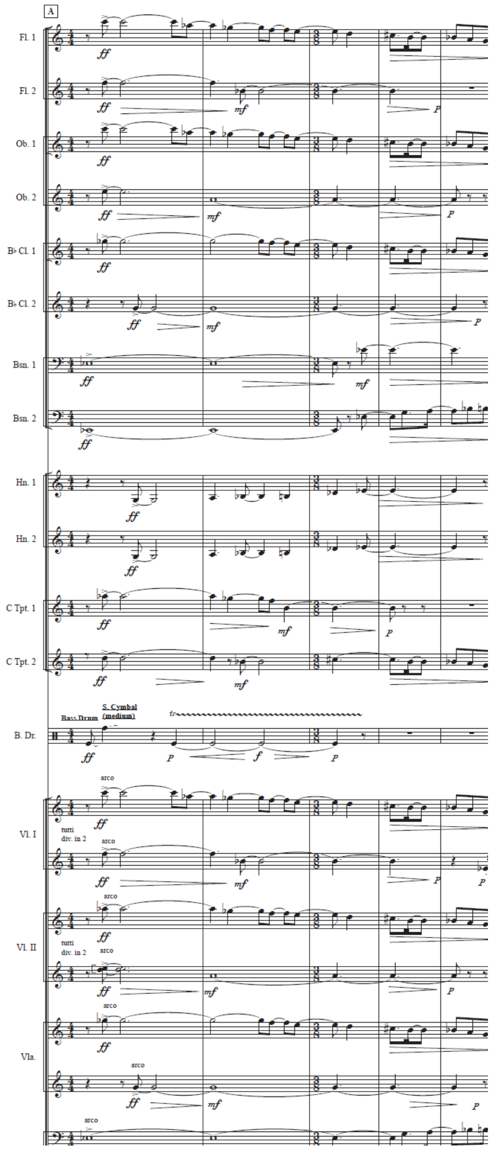
Critical band frequency: 345 Hz
Orchestration masking threshold: 52 dB
Target critical band excitation: 53 dB
Target to orchestration ratio: 1 dB
Target may be audible by trained ear on critical band

Example 87. Clicking the graph shows that the target may be audible to a trained ear on the current critical band.

The Expert view (the graph view in current the Score-Tool App version) at the point of the mouse click in Example 87 confirms the masking situation in the 3D graph. There are some

French horn peaks over the masking threshold, but most of the sound should be inaudible. To support this opinion, the spectral centroid of the orchestration seems to be much greater than the spectral centroid of the target. The French horn seems therefore to have a darker tone compared to the orchestration, and thus be even harder to hear in this passage (see Example 88). The variation coefficient is 0.506133, indicating a homogeneous orchestration. Based on the data, I assumed that the French horn would be masked, and the program suggests that the strongest masker would be the trumpet.



Example 89. Masking curve and spectral centroid bar-graph of the mouse click point of Example 87.

Example 88. Anonymous 2, work for chamber orchestra, rehearsal letter A, an orchestral *tutti*. The target instrument is the French horn (the 9th and 10th staves from the top). Both horns are playing in unison.

There was a slight misconception in what I told the composer and what was recorded in the composer's notes from the meeting. After the rehearsals, the composer reported to me on the third passage as follows:

> "Alun tutti. Se yksi trumpetti, joka oli paikalla tuli komeasti läpi, mutta käyrätorvet kyllä erotti mainiosti trumpetista huolimatta. kuten ohjelmasi arvioi. Ohjelma ennusti myös, että käyrätorvien väri olisi tummempi. Kenties, mutta kyllä soitin oli selkeästi edelleen tunnistettavissa."

> "Tutti at the beginning. The one trumpet present at the rehearsals came through gracefully, but the French horns could be heard fine despite the trumpet. Like your program estimated [*sic*]. The program estimated also that the sound color of the French horns would be darker. Probably so, but the instrument was clearly still recognizable."

This was an interesting case, as even though the dB rating of the French horn would rise by 3 because of the unison, most of the graph would still be under the masking threshold. I suspect the reason for the failed estimation could be the extra "brassy kick" in the sound that comes from moving from *forte* to *fortissimo*. An interesting test would be to play the same passage only *forte* and listen whether the result matches the estimated masking.

3.3.3 AFTERTHOUGHTS

The first field test showed that, as much as there is room for improvement, the program gives rational results. The biggest drawback is the user interface, which needs to be improved in order to be understandable by those not technically oriented. The composer in one of field tests was interested in the program but did not use the results directly to improve the orchestration. The feedback from the rehearsals was still valuable and revealed where the program needs to be improved.

The experience made me think that perhaps the optimal score for the program would be a solo concerto, where the audibility of the target instrument would be highly desirable, and unisons would not be such a problem. Still, among the described passages, two of the three predictions correlated with the hearing experience. In the current version of the Score-Tool App, the score can be virtually anything with notation symbols, because there is now, among other improvements, the possibility to select multiple targets.

EPILOGUE AND FUTURE RESEARCH

In this project I described the functions of the Score-Tool App in which I use the psychoacoustic model from the MPEG audio coding standard to determine the masking level of an orchestration against a target. In addition, I give an estimate of the target timbre distance to the orchestration, and target timbre brightness, i.e., the *spectral centroid* value, both of which have an impact on target audibility. The calculation accuracy increases if the orchestration and the target consist of one concurrently sounding chord and decrease if onset times of orchestration and target instruments are not simultaneous. In both cases, the Score-Tool App gives an estimate of target audibility, the *audibility prediction* value, which is valuable information for a composer who is composing an orchestral score and wants to try a new kind of orchestration.

The Score-Tool audibility prediction value covers only one aspect of everything that affects a target's audibility in a live performance situation. There are many parameters, especially psychological parameters, in an audience's listening experience that cannot be controlled or measured, such as the level of musical knowledge, the state of alertness, hearing disorder, focus of attention, and so on. With careful planning at the composition stage, it is, however, possible to ease the perception of the target as the "main thing" in a sounding image.

The goal of the project, namely, to develop an algorithm and an app to determine the target's audibility from the score, succeeded in part: the App gives reliable results about single orchestration chords. However, composers would also benefit from testing longer sections as a group, because many times the question is "Is this musical gesture audible?" rather than "Is this single note audible?" With the tools I introduced in this project, it would be difficult to determine the audibility of a musical gesture, so that would be a good research topic for the future. Also, the user interface of the Score-Tool App is currently too complicated for many of my colleagues to use. Thus, the user interface needs to be developed further, possibly with the help of more extensive user tests than were carried out in this project.

After collecting different audibility-related aspects during this project, I now summarize things to be considered while composing an orchestral work that includes a specific target:

How to make sure the target is audible in your score:

1. Use a target instrument with a unique timbre

An instrument's sound is considered audible when the instrument's characteristic timbre can be distinguished from the mass. Hence, a target's audibility can be enhanced by using a characteristic playing technique or *sordino* to alter the target's sound spectrum. The Score-Tool App's timbre distance parameter can be used to check how unique the target timbre is compared to the orchestration. If several instruments are chosen as targets, make sure they have matching timbres. Also, a target instrument timbre with rapid attack enhances audibility as well does the use of heavy vibrato.

2. Mark the target with a louder dynamic level than the orchestration

Soft dynamics blend better than loud dynamics. Sound is blended when an instrument's sound is still heard, not as a separate entity, but as a part of a complex timbre.

3. Use target with bright timbre

A low spectral centroid of the target timbre makes it easier to blend into the orchestration timbre.

4. Use the registral separation of the target if possible

Placing the target in its own register in the frequency space is a quite good way to ensure its audibility as long as there are no loud low-frequency timbres in the orchestration. The effect of registral separation can be confirmed with the Score-Tool App

5. Use the spatial separation on stage for conflicting instruments

The audibility of the solo instrument could be improved, if the soloist plays on far left or far right side of the stage. The effect of spatial separation can be tested in the Score-Tool App using the acoustic model of the Helsinki Music Centre Hall. On the other hand, visual focus on the soloist also improves audibility, so it is a good idea to place the target instrument on stage so that it attracts attention.

6. Make sure the target fulfills the criteria of "good continuation"

Avoid big leaps from one pitch to another because these break the "good continuation" and distract the listener from following the succession of notes.

7. Test the target's audibility in the Score-Tool App with "wrong" dynamics

Dynamics marked in a score are relative. The variance of realization is high, especially with soft dynamics. Because of the relativity of dynamics, it is good to test the audibility of the target with dynamics one step louder than indicated in the orchestration.

The Score-Tool App has also made an impact on my composition practice. Using the App, I feel comfortable trying something new in my orchestral works because I can test whether my new ideas would be audible in the actual performance. I use the App as a final check on my score after composing a full passage for orchestra, and I revise the score according to the results as needed. In this way the Score-Tool App resembles the proof-reading applications for written texts, which are often used as final check before submitting a text further.

The knowledge gained in this research project made me think of the concept of dynamics in a different light. The possibilities of interesting orchestral timbres are greater with soft dynamics than with loud dynamics. For example, two concurrent loud sounds together give the sensation of separated sound sources, whereas two soft sounds tend to blend easily as one source. Also, keeping the overall dynamic level of the orchestral work a step lower than my initial intention gives me room to experiment with the timbre balance in flexible way. My preference for soft dynamics can be seen, for example, in the score of my opera *All the Truths We Cannot See*. After composing the work, I noticed that in many passages I used the markings *mp* and *p* more than before. In the live performances I did not notice that the use of soft dynamics affected the intensity of the music.

3.4 FUTURE RESEARCH

In the future, one possible improvement on the audibility check would be to include onset times into the calculations. This would require taking psychological aspects into account, which would be an interesting research topic. The Score-Tool App engine also makes it possible to analyze the orchestration from other aspects besides audibility. For example, I have already implemented features that enable timbre classification possibilities for orchestral chords, the *timbre glyph* feature, which draws a comparable and classifiable image of the timbre. I have noticed in using the App that a full-bodied sounding orchestration has certain features that could be analytically pointed out, so that the full-bodiness (which I call *orchestration formant*) could be analyzed directly from the score. These features are already a part of the Score-Tool App, although they are not discussed within the scope of this report.

GLOSSARY OF TERMS

Algorithm	A sequence of clearly defined operations or instructions that can be implemented in a computer program. For example, the Pythagoras theorem is an algorithm used to solve the geometrics of an orthogonal triangle.
Amplitude	In acoustics, the height of a sound wave.
Auditory band	See **Critical band.**
Auditory stream	Connectedness of auditory events in a sequence. In auditory psychology, the sound or collection of sounds we focus on in a noisy environment. For example, in a cocktail party, our conversation partner's speech is most likely our auditory stream.
Bark	Approximation of critical bands in our hearing system: a list of fixed frequencies that have been used widely in acoustic calculations.
Beating	In acoustics, when two sinusoidal components are close to each other within one **critical band**, closer than a distance that causes **roughness**, the sensation of beating is perceived.
Binaural	The auditory sensation that involves listening to the sound with both ears.
Blending	In orchestration, the phenomenon in which two or more instrument sounds blend together in a listener's perception as one **super instrument**.
Brightness	See Spectral centroid.
Cepstrum	A play on the word spectrum by reversing the letters of the first syllable. Cepstrum is a special kind of inverse spectrum. The Cepstrum is the representation of periodic structures in the frequency spectrum.
Cocktail party effect	A term used for the human ability to concentrate on a single sound source in a noisy environment. For example, we are able to have a conversation with another person in a cocktail party, while hearing multiple other conversations at the same time.
Coefficient of variation	Used sometimes with acronym CV. In the Score-Tool App, this technique is used to estimate the homogeneity of the orchestration (**timbre homogeneity**). Applying this to the **MFCC** vector gives a value of how much the timbres of individual instruments deviate from

	the average of all orchestration timbres, i.e., it estimates the relative standard deviation of a given set of data.
Critical band	Width of an auditory filter in our hearing system. A frequency area where only the loudest frequency component can be perceived. Two equal-amplitude components give a mean frequency **roughness** or **beating** in the sound.
Critical distance	In a concert hall, the distance from a sound source beyond which the hall reverberations are louder than sound directly from the source. In concert halls, almost all seats are usually beyond the critical distance of the hall. For example, Boston Symphony Hall has a critical distance of 7 meters.
dB	Acronym for *deci Bel*, or *decibel*, a logarithmic and relative measure of sound pressure level. The actual unit is Bel, and deci means a tenth of a Bel. In acoustics, for example in concert hall, the sound pressure measured in pascals is referenced to the standard reference sound pressure in the air, which is 20 *micropascals*. The measured sound pressure in pascals is turned into decibels with the following formula: $20\log_{10}\left(\frac{measured\ pressure}{reference\ pressure}\right)dB$
Formant	A frequency area in which the amplitudes of the sound spectrum are emphasized due to resonance of vibrating body. Formant does not depend on pitch. For example, when a someone sings a scale using the vowel “a”, we identify the vowel as “a” because of vowels characteristic formants.
Harmonic	In music, an individual overtone of an instrument spectrum. The structure of harmonic overtones is always roughly the same: only the amplitudes of harmonics vary. If the structure of harmonics is distorted, as in a bell timbre, the sound is *inharmonic*.
Hearing system	In acoustics, the route that turns the received sound pressure waves into auditory sensations. Basically, a combination of outer, middle, and inner ear filters turns pressure waves into electricity, and the brain then interprets the electric potentials as auditory sensations.
Lossy audio coding	Method for compressing digital audio data so that it loses information in the process.
Loudness	Perceptive measure of sound pressure.
Masking	In psychoacoustics, the phenomenon whereby a sound is softer or inaudible because another louder sound is “taking up the space.”

Masking curve	In this project, the result of applying the MPEG psychoacoustic model to the orchestration sound spectrum, resulting in a curve indicating the level of masking on **critical bands** of our hearing system.
Mel scale	A scale resulting from a psychoacoustic test in which the subjects were asked to adjust the pitch of the test tone two times higher than the reference tone. 100 Mel is approximately 1 bark.[252]
MFCC	Acronym for Mel-Frequency Cepstrum Coefficients. A distilled version of the frequency spectrum that shows especially the formant areas.
Open source	In computer programming, a code that is made freely available to the general public. An open-source code is released under the terms of a license. Depending on the license terms, the code can be downloaded, modified, and published by anyone. All this can be done under the license terms of the Score-Tool App.
Orchestration chord	In the Score-Tool App, “everything that sounds together” in score, i.e., the combination of instruments playing at a specific point in a score.
Panning	Moving the sound source in stereo image in-between left and right.
Partial	A sine wave component of a complex sound. A partial can be a harmonic- or an inharmonic partial.
Psychoacoustic model	Computational model of the human hearing system that takes into account the audio perception, such as the effect of critical bands.
Register	In orchestration, an ambiguous term for a particular segment in instrument’s range or a region of the voice set off by vocal breaks. Some instruments have names for registers: for example, clarinet has the chalumeau (low), clarion (middle), and altissimo (high) registers, caused by physical properties of the instrument. In some instruments, such as the organ and accordion, register refers to timbre.
Roughness	In acoustics, when two sinusoidal components are close to each other within one critical band, further than a distance that causes beating, the sensation of roughness is perceived. Roughness can be generated by any amplitude modulation of sufficient depth and rate.
Singer’s formant	A characteristic of the voice of a trained opera singer. Sometimes the singer’s formant phenomenon is claimed to be the cause of the

---

[252] Pulkki and Karjalainen 2015, p. 174.

	audibility of the singer’s voice when a *tutti* orchestra is playing in the background.
Sordino	A mechanical device that acts as a filter and alters an instrument’s timbre.
Sound color	See Timbre.
Spectral dominant region	The frequency area around 700 Hz; the area that gets the most attention in our hearing system. This area is, for example, essential for speech intelligibility.
Spectral centroid	The center of mass of the sound. For example, if an instrument sound has very a strong base frequency, the spectral centroid would be close to the notated pitch. If an instrument sound has a rich harmonic content, the spectral centroid is much higher than the notated pitch. The spectral centroid is sometimes used as a measure of the brightness of a sound.
Spreading function	In masking, a function that defines how much the masking effect in one critical band leaks into adjacent bands.
SPL	Acronym for Sound Pressure Level.
Squillo	See Singer’s formant.
Super instrument	In orchestration, a combination of instruments playing together, resulting in a sound that is perceived as coming from one source.
Target	In this report, the sound of an instrument or an instrumental group whose audibility is tested against the orchestration.
Texture	In orchestration, a compositional element that a composer perhaps intended to sound as one uniform entity. An example of a texture could be a long steady tone of horn to which is added a repetitive clarinet triad arpeggio and a long violin trill.
Timbre	Property of sound that is not related to pitch, dynamics, duration, or spatial position. Sometimes referred to as sound color.
Timbre homogeneity	In orchestration, when the timbre consists of two or more concurrent sounding instrument sounds, the homogeneity is the degree of similarity among constituent timbres.

Virtual pitch	A pitch that has no energy at the fundamental frequency. The virtual pitch is the best candidate for the fundamental that fits into actual sounding overtones.

REFERENCES

Abdi, H. 2010. “Coefficient of Variation.” In N.J. Salkind, D.M., Dougherty, & B. Frey (Eds.): Encyclopedia of Research Design. Thousand Oaks (CA): Sage. pp. 169-171.

ACTOR Project website. 2020. “ACTOR Project.” June 11, 2020. https://www.actorproject.org/.

Adler, Samuel. 2016. *The Study of Orchestration*. Fourth edition. New York: W. W. Norton & Company.

Alam, Mahbubul. 2020. “Learning Data Visualization Differently.” Medium. November 25, 2020. https://towardsdatascience.com/learning-data-visualization-differently-87204b8e3216.

Ashihara, Kaoru. 2007. “Hearing Thresholds for Pure Tones above 16kHz.” *The Journal of the Acoustical Society of America* 122 (3): EL52–57. https://doi.org/10.1121/1.2761883.

Assayag, Gérard, Camilo Rueda, Mikael Laurson, Carlos Agon, and Olivier Delerue. 1999. “Computer-Assisted Composition at IRCAM: From PatchWork to OpenMusic.” *Computer Music Journal* 23 (3): 59–72.

Barron, Michael. 2009. *Auditorium Acoustics and Architectural Design*. 2nd edition. London ; New York: Taylor & Francis.

Benediktsson, Björgvin. 2019. *Read Step By Step Mixing Online by Björgvin Benediktsson | Books*. https://www.scribd.com/book/507015510/Step-By-Step-Mixing-How-to-Create-Great-Mixes-Using-Only-5-Plug-ins.

Berlioz, Hector, and Richard Strauss. 1948. *Treatise on Instrumentation (Including Berlioz ’ Essay on Conducting (1844, 1855, 1904)*. Edwin F. Kalmus.

Bigand, Emmanuel, Charles Delbé, Yannick Gérard, and Barbara Tillmann. 2011. “Categorization of Extremely Brief Auditory Stimuli: Domain-Specific or Domain-General Processes?” *PLOS ONE* 6 (10): e27024. https://doi.org/10.1371/journal.pone.0027024.

Bindu, K. Hima, Raghava Morusupalli, Nilanjan Dey, and C. Raghavendra Rao. 2021. *Coefficient of Variation and Machine Learning Applications*. 1st edition. CRC Press. https://doi.org/10.1201/9780429296185.

Blauert, Jens. 1974. *Räumliches Hören, 3 Bde.* Hirzel, Stuttgart.

Carpentier, Grégoire, Damien Tardieu, Jonathan Harvey, Gérard Assayag, and Emmanuel Saint-James. 2010. “Predicting Timbre Features of Instrument Sound Combinations: Application to Automatic Orchestration.” *Journal of New Music Research* 39 (1): 47–61. https://doi.org/10.1080/09298210903581566.

Chen, Chun-houh, Wolfgang Karl Härdle, and Antony Unwin, eds. 2007. *Handbook of Data Visualization*. 2008th edition. Berlin: Springer.

Chiasson, Frédéric, Caroline Traube, Clément Lagarrigue, and Stephen McAdams. 2017. “Koechlin’s Volume: Perception of Sound Extensity among Instrument Timbres from Different Families.” *Musicae Scientiae* 21 (1): 113–31. https://doi.org/10.1177/1029864916649638.

Chon, Song Hui, and David Huron. 2014. “DOES AUDITORY MASKING EXPLAIN THE HIGH VOICE SUPERIORITY?” *ICMPC,* 4, School of Music Ohio State University.

Concerthalls.org website. 2021. “Concert Hall Reverberation.” *Concerthalls.Org* (blog). 2021. http://www.concerthalls.org/?page_id=84.

Daigle, Alexandre. 2020. “Evaluation of Optical Music Recognition Software.” McGill University. https://escholarship.mcgill.ca/concern/theses/zc77sv75m.

Davis, S., and P. Mermelstein. 1980. “Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences.” *IEEE Transactions on Acoustics, Speech, and Signal Processing* 28 (4): 357–66. https://doi.org/10.1109/TASSP.1980.1163420.

Debussy, Claude. 1900. *Nocturnes*. Paris, France: E. Fromont.

Deutsch, Diana. 2012. *The Psychology of Music*. 3rd ed. Amsterdam: Academic Press.

Goodchild, Meghan & Mcadams, Stephen. 2018. Perceptual Processes in Orchestration. in *The Oxford Handbook of Timbre (2021)*. New York: Oxford University Press. 10.1093/oxfordhb/9780190637224.013.10.

DPA Microphones. 2021. “Acoustical Characteristics of Musical Instruments.” DPA. September 19, 2021. https://www.dpamicrophones.com/mic-university/acoustical-characteristics-of-musical-instruments.

Eastern Kentucky University website. 2021. “Health And Safety Information | School Of Music | Eastern Kentucky University.” September 19, 2021. https://music.eku.edu/health-and-safety-information.

Everitt, B. S., and Anders Skrondal. 2010. *The Cambridge Dictionary of Statistics*. 4th edition. Cambridge, UK ; New York: Cambridge University Press.

Fastl, Hugo, and Eberhard Zwicker. 2007. *Psychoacoustics: Facts and Models*. 3rd ed. Berlin Heidelberg: Springer-Verlag. https://doi.org/10.1007/978-3-540-68888-4.

Fletcher, Harvey, and W. A. Munson. 1933. “Loudness, Its Definition, Measurement and Calculation*.” *Bell System Technical Journal* 12 (4): 377–430. https://doi.org/10.1002/j.1538-7305.1933.tb00403.x.

French, N. R., and J. C. Steinberg. 1947. “Factors Governing the Intelligibility of Speech Sounds.” *The Journal of the Acoustical Society of America* 19 (1): 90–119. https://doi.org/10.1121/1.1916407.

Fuhrmann, Ferdinand. 2012. “Automatic Musical Instrument Recognition from Polyphonic Music Audio Signals.” PhD Thesis, Universitat Pompeu Fabra. http://hdl.handle.net/10803/81328.

Fuhrmann, Ferdinand, Martin Haro, and Perfecto Herrera. 2009. *Scalability, Generality and Temporal Aspects in Automatic Recognition of Predominant Musical Instruments in Polyphonic Music.* Conference: Proceedings of the 10th International Society for Music Information Retrieval Conference, ISMIR 2009, Kobe International Conference Center, Kobe, Japan, October 26-30, pp. 321-326.

Gelfand, Stanley A. 2009. *Hearing: An Introduction to Psychological and Physiological Acoustics*. 5th edition. New York: Informa Healthcare.

Griesinger, David. 2011. “THE RELATIONSHIP BETWEEN AUDIENCE ENGAGEMENT AND THE ABILITY TO PERCEIVE PITCH, TIMBRE, AZIMUTH AND ENVELOPMENT OF MULTIPLE SOURCES.” *Proceedings of the Institute of Acoustics*, May 2011.

Han, Yoonchang, Jaehun Kim, and Kyogu Lee. 2017. “Deep Convolutional Neural Networks for Predominant Instrument Recognition in Polyphonic Music.” *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 25 (1): 208–21. https://doi.org/10.1109/TASLP.2016.2632307.

Harvey, Jonathan. 2008. *Speakings*. https://www.fabermusic.com/music/speakings-5282.

Huron, David. 1989. “Voice Denumerability in Polyphonic Music of Homogeneous Timbres.” *Music Perception* 6 (4): 361–82. https://doi.org/10.2307/40285438.

Huron, David. 1993. “Note-Onset Asynchrony in J. S. Bach’s Two-Part Inventions.” *Music Perception: An Interdisciplinary Journal* 10 (July): 435–43. https://doi.org/10.2307/40285582.

Inácio, Octávio. 2005. “FUNDAMENTALS OF ROOM ACOUSTICS.” *IACMA - International Advanced Course on Musical Acoustics*, no. Bologna, Italy: 31.

ISO/IEC. 1996. “ISO/IEC 13818-1:1996 Generic Coding of Moving Pictures and Associated Audio Information.” https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/02/29/22989.html.

Järveläinen, Hanna, Vesa Välimäki, and Matti Karjalainen. 1999. “Audibility of Inharmonicity in String Instrument Sounds, and Implications to Digital Sound Synthesis.” in String Instrument Sounds, and Implications to Digital Sound Synthesis.” International Conference on Mathematics and Computing.

Juth, Thomas. 2021. *The Art of Equalization*. Kindle edition. https://www.goodreads.com/book/show/57757002-the-art-of-equalization

Kidd, Gerald Jr., Christine R. Mason, Virginia M. Richards, Frederick J. Gallun, and Nathaniel I. Durlach. 2008. *Auditory Perception of Sound Sources*. Springer Handbook of Auditory Research 29. New York: Springer.

Kim, Duck Hoon, In Park, Il Yun, and Sang Lee. 2004. “A New MPEG-7 Standard: Perceptual 3-D Shape Descriptor.” In , 3332:238–45. https://doi.org/10.1007/978-3-540-30542-2_30.

Kirby, David, and Kaoru Watanabe. 1996. “Report on the Formal Subjective Listening Tests of MPEG-2 NBC Multichannel Audio Coding. | MPEG.” *MPEG 37 - Maceió*, no. November 1996. https://mpeg.chiariglione.org/standards/mpeg-2/advanced-audio-coding/report-formal-subjective-listening-tests-mpeg-2-nbc.

Kitchen, Nicholas. 2017. “Beethoven-Op.-131-Manuscript-Markings.” In . https://www.bu.edu/beethovencenter/files/2017/09/Beethoven-Op.-131-manuscript-markings.pdf.

Kleczkowski, Piotr, M Plewa, and Marek Pluta. 2011. “Masking a Frequency Band in a Musical Fragment Played by a Single Instrument.” *Acta Physica Polonica Series a* 119 (June). https://doi.org/10.12693/APhysPolA.119.991.

Krämer, Johannes. 2011. “Sound Power of Modern and Historical Orchestral Instruments.” INSTITUT FÜR SPRACHE UND KOMMUNIKATION FACHGEBIET MUSIKWISSENSCHAFT: Technishche Universität Berlin.

Kuttruff, Heinrich. 2006. *Acoustics: An Introduction*. 1st edition. London ; New York: CRC Press.

Laurson, Mikael. 1996. *PatchWork: A Visual Programming Language and Some Musical Applications*. Sibelius Academy.

Laurson, Mikael, Mika Kuuskankare, and Vesa Norilo. 2009. “An Overview of PWGL, a Visual Programming Environment for Music.” *Computer Music Journal* 33 (1): 19–31. https://doi.org/10.1162/comj.2009.33.1.19.

Lembke, Sven-Amin. 2015. *When Timbre Blends Musically: Perception and Acoustics Underlying Orchestration and Performance*. McGill University Libraries.

Lembke, Sven-Amin, and Stephen McAdams. 2015. “The Role of Spectral-Envelope Characteristics in Perceptual Blending of Wind-Instrument Sounds.” *Acta Acustica United with Acustica* 101 (5): 1039–51. https://doi.org/10.3813/AAA.918898.

Lincoln, Bo. 1998. “An Experimental High Fidelity Perceptual Audio Coder.” 1998. https://www.semanticscholar.org/paper/An-Experimental-High-Fidelity-Perceptual-Audio-Lincoln/69e00b7a6d6bc9b59b42a4a58bf2c30e925774ea.

McAdams, Stephen. 2019. “Timbre as a Structuring Force in Music.” In *Timbre: Acoustics, Perception, and Cognition*, edited by Kai Siedenburg, Charalampos Saitis, Stephen McAdams, Arthur N. Popper, and Richard R. Fay, 211–43. Springer Handbook of Auditory Research. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-14832-4_8.

Mcadams, Stephen, and Albert Bregman. 1979. “Hearing Musical Streams.” *Computer Music Journal* 3 (January).

Meyer, Jürgen. 2009. *Acoustics and the Performance of Music: Manual for Acousticians, Audio Engineers, Musicians, Architects and Musical Instrument Makers*. 5th ed. Modern Acoustics and Signal Processing. New York: Springer-Verlag. https://doi.org/10.1007/978-0-387-09517-2.

Middlebrooks, John C., and Jonathan Z. Simon. 2017. “Ear and Brain Mechanisms for Parsing the Auditory Scene.” In *The Auditory System at the Cocktail Party*, edited by John C. Middlebrooks, Jonathan Z. Simon, Arthur N. Popper, and Richard R. Fay, 1–6. Springer Handbook of Auditory Research. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-51662-2_1.

Miśkiewicz, Andrzej, and Andrzej Rakowski. 1994. “Loudness Level versus Sound-pressure Level: A Comparison of Musical Instruments.” *The Journal of the Acoustical Society of America* 96 (6): 3375–79. https://doi.org/10.1121/1.411448.

Moore, Brian C. J. 2012. *An Introduction to the Psychology of Hearing*. BRILL.

Noh, Ohran. 2009. “Wolfgang Amadeus Mozart’s Piano Sonata in F Major, KV 533/494 in the Arrangement by Edvard Grieg : A Critical Examination of the Musical Text in the Context of the Primary Sources,” August. https://harvest.usask.ca/handle/10388/etd-09212009-134335.

Noll, P. 1997. “MPEG Digital Audio Coding.” *IEEE Signal Processing Magazine* 14 (5): 59–81. https://doi.org/10.1109/79.618009.

Orchidea website. 2020. “Orchidea.” June 11, 2020. http://www.orch-idea.org/.

Pätynen, Jukka. 2011. *A Virtual Symphony Orchestra for Studies on Concert Hall Acoustics*. Aalto University Publication Series DOCTORAL DISSERTATIONS , 86/2011. Department of Media Technology: Aalto University. https://aaltodoc.aalto.fi:443/handle/123456789/5055.

Pätynen, Jukka, and Tapio Lokki. 2010. “Directivities of Symphony Orchestra Instruments.” *Acta Acustica United with Acustica* 96 (1): 138–67. https://doi.org/10.3813/AAA.918265.

Pätynen, Jukka, and Tapio Lokki. 2016. “Perception of Music Dynamics in Concert Hall Acoustics.” *The Journal of the Acoustical Society of America* 140 (5): 3787–98. https://doi.org/10.1121/1.4967157.

Pedrielli, Francesca, Eleonora Carletti, and Camilla Casazza. 2008. “Just Noticeable Differences of Loudness and Sharpness for Earth Moving Machines.” *The Journal of the Acoustical Society of America* 123 (5): 3164–3164. https://doi.org/10.1121/1.2933219.

Peeters, Geoffroy. 2004. “A Large Set of Audio Features for Sound Description (Similarity and Classification) in the CUIDADO Project.” *CUIDADO Ist Project Report* 54 (January): 1–25.

Pepinsky, Abe. 1941. “Masking Effects in Practical Instrumentation and Orchestration.” *The Journal of the Acoustical Society of America* 12 (3): 405–8. https://doi.org/10.1121/1.1916120.

Piston, Walter. 1955. *Orchestration*. 1st edition. New York: W. W. Norton & Company.

Plazak, Joseph, David Huron, and Benjamin Williams. 2010. “Fixed Average Spectra of Orchestral Instrument Tones.” *Empirical Musicology Review* 5 (1): 10–17. https://doi.org/10.18061/1811/45171.

Pulkki, Ville, and Matti Karjalainen. 2015. *Communication Acoustics: An Introduction to Speech, Audio and Psychoacoustics*. 1st edition. Chichester, West Sussex, United Kingdom: Wiley.

Ravel, Maurice. 1929. *Boléro*. Paris, France: DURAND.

Reuter, Christoph. 1996. *Die auditive Diskrimination von Orchesterinstrumenten: Verschmelzung und Heraushörbarkeit von Instrumentalklangfarben im Ensemblespiel*. Frankfurt am Main ; New York: Peter Lang GmbH, Internationaler Verlag der Wissenschaften.

Reuter, Christoph. 2000. “Blending and partial masking.” *Proccedings of the Blagodatovskije readings*, December, S. 180-183.

Reuter, Christoph. 2004. “Vibrato and Micromodulations as Cues for Timbre Segregation.” *Proceedings of the Russian Institute of History of Arts and European Organology*, no. December 6-10 (Petersburg): S. 168-171.

Reuter, Christoph. 2014. “Zusammenklang von Musikinstrumenten.” *Weinzierl, Stefan (Hrsg.): Akustische Grundlagen Der Musik*, no. Laaber (Laaber): S. 391-408.

Riddle, Nelson. 1985. *Arranged by Nelson Riddle*. Alfred Music.

Rimsky-Korsakov, Nikolay. 2013. *Principles of Orchestration*. CreateSpace Independent Publishing Platform.

Sandell, Gregory John. 1991. “Concurrent Timbres in Orchestration: A Perceptual Study of Factors Determining ‘Blend.’” *Ph.D. Thesis*. https://ui.adsabs.harvard.edu/abs/1991PhDT.......313S.

Schönberg, Arnold. 1912. *Fünf Orchesterstücke*. Vol. Opus 16. Leibzig, Germany: Edition Peters.

Shinn-Cunningham, Barbara, Virginia Best, and Adrian K. C. Lee. 2017. "Auditory Object Formation and Selection." In *The Auditory System at the Cocktail Party*, edited by John C. Middlebrooks, Jonathan Z. Simon, Arthur N. Popper, and Richard R. Fay, 7–40. Springer Handbook of Auditory Research. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-51662-2_2.

Silverman, Robert. 2020. "Silverman on Mozart." December 23, 2020. https://www.forte-piano-pianissimo.com/SilvermanonMozart.html.

Smith, J.O., and J.S. Abel. 1999. "Bark and ERB Bilinear Transforms." *IEEE Transactions on Speech and Audio Processing* 7 (6): 697–708. https://doi.org/10.1109/89.799695.

Sørensen, Jesper B. 2002. "The Use and Misuse of the Coefficient of Variation in Organizational Demography Research." *Sociological Methods & Research* 30 (4): 475–91. https://doi.org/10.1177/0049124102030004001.

Stephen Downie. 2003. "Music Information Retrieval." In *Music Information Retrieval (Chapter 7). In Annual Review of Information Science and Technology 37*, 295–340. 37. Medford, NJ: Information Today, 2003.

Stöter, Fabian-Robert, Michael Schoeffler, Bernd Edler, and Jürgen Herre. 2013. "Human Ability of Counting the Number of Instruments in Polyphonic Music." *Proceedings of Meetings on Acoustics* 19 (1): 035034. https://doi.org/10.1121/1.4799609.

Sundberg, J. 1977. "The Acoustics of the Singing Voice." *Scientific American* 236 (3): 82–84, 86, 88–91. https://doi.org/10.1038/scientificamerican0377-82.

Tardieu, Damien, and Stephen McAdams. 2012. "Perception of Dyads of Impulsive and Sustained Instrument Sounds." *Music Perception: An Interdisciplinary Journal* 30 (2): 117–28. https://doi.org/10.1525/mp.2012.30.2.117.

Temppeliaukionkirkko website. 2021. "Temppeliaukionkirkko." 2021. https://www.temppeliaukionkirkko.fi/.

Terhardt, Ernst. 1979. "Calculating Virtual Pitch." *Hearing Research* 1 (2): 155–82. https://doi.org/10.1016/0378-5955(79)90025-X.

Terhardt, Ernst, Gerhard Stoll, and Manfred Seewann. 1982. "Algorithm for Extraction of Pitch and Pitch Salience from Complex Tonal Signals." *The Journal of the Acoustical Society of America* 71 (3): 679–88. https://doi.org/10.1121/1.387544.

Todoroff, Todor; Daubresse. 1995. "IANA: A Real-Time Environment for Analysis and Extraction of Frequency Components of Complex Orchestral Sounds and Its Application within a Musical Realization." *International Computer Music Conference Proceedings* 1995. http://hdl.handle.net/2027/spo.bbp2372.1995.088.

Wakefield, Jonathan P., and Christopher Dewey. 2015. "An Investigation into the Efficacy of Methods Commonly Employed by Mix Engineers to Reduce Frequency Masking in the Mixing of Multitrack Musical Recordings: 138th Audio Engineering Society Convention." *138th Audio Engineering Society Convention, AES 2015* 2: 1044–49.

Wegman, Edward, and Jeffrey Solka. 2004. "Statistical Data Mining." *Handbook of Statistics* 24 (January): 1–46. https://doi.org/10.1016/S0169-7161(04)24001-9.

Wilke, Claus O. 2019. *Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures*. 1st edition. Sebastopol, CA: O'Reilly Media.