

Multidimensional Timbre Space Mapping via the Spherical Harmonic Transform

Everett M. Carpenter

October 7, 2026

1 Introduction

In computer music composition, the vast range of possibilities can oftentimes be overwhelming to the composer or user. Thus, tools to facilitate effective usage of compositional parameters must be developed. Examples of these can be seen in software such as the graphical languages of Max MSP, and Pure-Data, the text-based languages of SuperCollider, CSound, and ChucK, as well as the commercial software of DAWs and VSTs. All aforementioned software offer an abstraction of the sound processing that the computer provides. One may find an explanation of such an abstraction by a quick look at the "Audio Buffer". The audio buffer provides a list of sample values which when played back, constitute a small segment of sound. Through operations such as addition, subtraction, multiplication, and division, the user can apply well known effects such as tremolo, vibrato, time stretching, filtering, and many more. While the enormous quantity of possibilities may be appealing, the minuscule scale and high level of direct involvement with sample values do not lend themselves well to those interested in editing sound. Therefore the software must create an abstraction containing a *reduction of parameters*.

Timbre has proven to be an effective parameter in musical composition, and equips the artist with a broad brush to paint their sonic landscape. However, timbre itself is a collection of parameters; it requires its own reduction in dimensionality. Even more so, a concept of timbre is oftentimes fueled by a collection of perception metrics which can be both causal and non-causal. Thus timbre is a parameter subject to definition by the user. Research and development in the effectiveness of tools for timbre selection has persisted for many decades.¹ Research has investigated the organization of timbres into *timbre spaces*, and the control of these spaces as effective methods for composition and performance^{2,3,4}. Due to the large number of dimensions that these spaces can fall

1. Paul E. Dworak and Philip C. Baczewski, "A Vector Field Model of Compositional Creativity" (January 1980); Reynald Hoskinson and Dinesh Pai, "Manipulation and Resynthesis with Natural Grains" (January 2001).

2. David Creasey, David Howard, and Andrew Tyrrell, "The Timbral Object - An Alternative Route to the Control of Timbre Space" (January 1996).

3. Joshua Ryan Lam and Charalampos Saitis, "The Timbre Explorer: A Synthesizer Interface for Educational Purposes and Perceptual Studies," *International Conference on New Interfaces for Musical Expression*, April 2021,

4. Jeff Gregorio and Youngmoo E. Kim, "Evaluation of Timbre-Based Control of a Parametric Synthesizer," <https://nime.pubpub.org/pub/adtb2zl5>, *International Conference on New*

within, conventional three or two dimensional vector visualizations do not effectively convey the timbre actively presented. Two solutions to the visualization as well as control of timbre spaces will be presented in the next section.

2 Timbre Space Realization

As mentioned previously, organization and control of timbre spaces has been the key interest in research which centers on the technology. The former focuses largely on intuitive and meaningful methods of categorizing timbres, and associating them within a dataset which reveals their similarity measures. Criteria for categorization include perception metrics which center around auditory descriptors^{5, 6}. These descriptors are helpful in evaluating the similarities between sounds, but timbre still remains a subjective term. A metric-based process provides a solid skeletal map for a timbre space, but final changes must be made by users using their own subjective input.

2.1 Perception Metrics as a Basis

Metric-based timbre mapping looks to provide a map of sounds whose creation requires little to no manual input. However, it has been established that timbre is a quality determined by the relationship between subject and object. We must then develop a system which models the most objective components of timbre perception,⁷ and provide a flexible interface for components which are indeed subjective. Terasawa et al. show that a statistical model for the perceived *distance* between two timbres may be derived.⁸ Therefore we may utilize the extensively researched methods of the human auditory system and model its ability to analyze a collection of sounds which are correlated based upon the results of this analysis, thus producing a large data set (frequently referred to as a *corpus*) which mirrors that used by humans during their subjective process of timbre perception. Oftentimes, listeners are capable of distinguishing sounds without being aware of the spectral and temporal measures their auditory systems are conducting. Nevertheless, they are provided an incredibly detailed dataset whose accuracy only depends upon mental and physical health. It is through the process of learning and understanding these datasets that the listener develops skills to measure distance between sounds across many dimensions. The intertwinement of auditory cognition and observation is effectively modelled

Interfaces for Musical Expression, April 2021,

5. Séverine Samson, Robert Zatorre, and James Ramsay, “Multidimensional Scaling of Synthetic Musical Timbre: Perception of Spectral and Temporal Characteristics,” *Canadian journal of experimental psychology = Revue canadienne de psychologie expérimentale* 51 (January 1998): 307–15, <https://doi.org/10.1121/1.406009>.

6. John M. Grey, “Multidimensional Perceptual Scaling of Musical Timbres,” *The Journal of the Acoustical Society of America* 61, no. 5 (May 1977): 1270–1277, ISSN: 0001-4966, <https://doi.org/10.1121/1.381428>, eprint: https://pubs.aip.org/asa/jasa/article-pdf/61/5/1270/11635616/1270_1_online.pdf, <https://doi.org/10.1121/1.381428>.

7. Hiroko Terasawa, Malcolm Slaney, and Jonathan Berger, “The Thirteen Colors of Timbre” (November 2005), 323–326, ISBN: 0-7803-9154-3, <https://doi.org/10.1109/ASPAA.2005.1540234>.

8. Hiroko Terasawa, Malcolm Slaney, and Jonathan Berger, “A Statistical Model of Timbre Perception,” in *ITRW on Statistical and Perceptual Audio Processing (SAPA 2006)* (2006), 18–23.

through a collection of *metric spaces*,⁹ where distance metrics—auditory cognitive function—are defined and applied to sets—auditory observation results—for interpretation.

Various tools combining statistical methods and psychoacoustic factors have been developed as a procedure for analyzing sound and its perception by the auditory system. Two of the four metrics used in *The Timbre Explorer*¹⁰ are rooted in these measures, both spectral *flux* and *density*. Both of these qualities are informed by the ear’s ability to perform a lossy *Fourier Transform*. For most listeners, spectral flux and density are familiar auditory phenomena, but are often felt more than articulated. These concepts already provide a reduction of dimensionality in themselves, as they are most significantly temporal; thus they are effective parameters to provide the user. The listener’s ability to observe spectral and temporal characteristics of sound and then correlate them will inform the design of interfaces for timbre mapping.

2.2 Timbre Space Creation as a Subjective Process

A detailed dataset requires intuitive controls to assist interpretation and navigation of said set. Geometric models have been introduced,¹¹ utilizing the common notion of Euclidean space while associating colors with timbres. An image of Wessel’s *space-master* can be viewed in Fig. 1. Wessel’s model (built in Max MSP) looks to interpolate between lists, effectively creating a parameter space. These lists are then utilized as parameters for sound synthesis, sample playback, reverberation models, and rhythmic patterns. An important feature of Wessel’s *space-master* is that it allows the space to be completely user-defined. Its construction is informed solely by the user’s conscious—or subconscious—interpretation of auditory data. While not explicitly employing perception metrics, the space created (unless done completely at random) is informed by the user’s perception. As previously discussed, the user’s cognitive mapping of a collection of sounds is analogous to a collection of metric spaces. Thus, Wessel’s model is not only perceptually fueled, but is also subjective.

3 Spherical Harmonic Projections as Spatial Encoding

In ambisonics, a collection of spherical harmonics is used as a basis onto which positional vectors are projected. This projection is the fundamental operation of ambisonics; it is used to map a position of continuous spherical coordinates to a finite space which is made up of multiple basis vectors. The process of projection attributes positional data to sound, encapsulated within a multichannel audio stream. Each channel of the audio stream represents the projection of the

9. B. Mendelson, *Introduction to Topology: Third Edition*, Dover Books on Mathematics (Dover Publications, 2012), ISBN: 9780486135090, <https://books.google.com/books?id=FWFmoEUJSwkC>.

10. Lam and Saitis, “The Timbre Explorer: A Synthesizer Interface for Educational Purposes and Perceptual Studies.”

11. Ali Momeni and David Wessel, “Characterizing and Controlling Musical Material Intuitively with Geometric Models,” in *Proceedings of the International Conference on New Interfaces for Musical Expression*, Montreal, Canada (Zenodo, June 2003), 54–62, <https://doi.org/10.5281/zenodo.1176535>, <https://doi.org/10.5281/zenodo.1176535>.

audio upon a spherical harmonic. Due to the orthonormality of the spherical harmonics, multiple streams of audio may be mapped independently and subsequently combined, all the while preserving their independent mappings. Such a preservation has enabled the process to be a viable option for spatial audio transmission and creation. The process of ambisonic encoding can be expressed mathematically as,

$$\vec{\beta}(t) = s(t) \cdot \vec{y}(\theta_s, \phi_s) \quad (1)$$

Where $\vec{\beta}$ is a vector containing the resulting multichannel audio stream describing the sphere at time t , $\vec{y}$ is a column vector containing the spherical harmonic projections of the source material’s ideal position (θ_s, ϕ_s) , and $s(t)$ is the source audio stream. This process acts as a frequency-independent spatial mapping or encoding, which is well-defined in ambisonics literature. A multichannel stream derived from this method (from now on referred to as a *spherical signal*) can be subsequently decoded via the following calculation, commonly referred to as sampling ambisonic decoding (SAD),¹²

$$\alpha(t) = \vec{\beta}(t) \cdot \vec{g}(\theta_p, \phi_p) \quad (2)$$

Here, α is a monophonic signal describing the sound pressure located at (θ_p, ϕ_p) , $\vec{\beta}$ is the aforementioned multichannel audio stream, and $\vec{g}$ is a vector containing the spherical harmonic projections of the ideal positioning (θ_p, ϕ_p) . Such a process can be described as a frequency-independent *beamforming*, although ambisonic sampling decoding for loudspeakers follows the exact same formulation. Thus, $\vec{\beta}$ is an intermediate space which describes an object or collection of objects via their audio contents and spatial position upon the sphere. How we choose to *listen* to this signal is where creative applications of technology may arise.

4 Spherical Harmonic Projections in Timbre Mapping

As mentioned previously, ambisonic processing has been developed with the application of spatial audio in mind, focusing on reconstructing the spatial image for headphones and speaker arrays. Its consideration in the field of sound synthesis has been virtually non-existent. Mapping continuous spaces to finite ones is a rather compelling concept in regard to artistic creativity, and poses a new view on attributing and interpreting data. To explore this potential, the process of ambisonic encoding and decoding has been recontextualized as a method for timbre space creation called *Spherical Timbre Organization* (STO). STO allows the spatial encoding of sounds or parameters into a virtual spherical space, which can be decoded in a number of ways. Due to the nature of this recontextualization, nearly all operations in the ambisonic domain apply to STO including spatial warping, rotation, mirroring, and distortions.¹³

12. Matthias Frank. and Frank Zotter, *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality* (Creative Media Partners, LLC, 2020), ISBN: 9781013273193, <https://books.google.com/books?id=LET2zQEACAAJ>.

13. Matthias Kronlachner, “Spatial Transformations for the Alteration of Ambisonic Recordings” (master’s thesis, Institut für Elektronische Musik und Akustik, 2013).

The initial workflow developed for STO contained three stages, where two types of audio sources were encoded: looping audio files and real-time additive synthesis algorithms. The process was carried out in the following order,

1. Positioning audio sources in space.
2. Visualizing the sound sphere.
3. Beamform decoding of the sound sphere.

The first step is a subjective process which allows the user to position sounds within a space solely designed by themselves. In the initial software design, this was accomplished by the calculation of spherical harmonic projections via reduced spherical coordinates (only using azimuth θ and elevation ϕ). Using audio, multiple streams of the audio to be encoded are weighted with gains corresponding to the spherical harmonic projections. For control data, the same is done but only when the harmonics projections are calculated, as the encoding does not need to be done at sample rate. Philosophically, it is thought of the same as the construction of a space utilizing ambisonics for the design and sounding of a virtual space. Thus, the design of a timbre space via spherical harmonic projections can be thought of as the design of a *timbre room* or *environment*. Various artistic metaphors can be derived from this concept; consider the phenomenon of noise pollution in the sonic environment of the world, and its translation to a timbre environment. Through various spatial distortions, one could modify the timbre environment to be unrecognizable to the performer. This renders the environment itself as a parameter for sound creation, and not just the method used to hear it.

The intermediate step which is concerned with the visualization of sonic contents within the sphere can be viewed in Fig. 2. While simple, it conveys the spatial positioning of the audio sources. Further expansion of this visualization could include a Gaussian kernel representing the current amplitude value of a source (as used in Wessel's demonstration in Fig. 3). Additionally, perception metrics such as spectral centroid, kurtosis or flatness could be viable visualizations for the facilitation of timbre space navigation. Associating color to spectral features could be a fruitful method of displaying where significantly different timbres may lie. A more straightforward visualization could be associating the sound's size to the current root mean square (RMS) value at some point in time. By displaying a set of sources and their positions relative to each other, the user can easily navigate the space with an awareness of their surroundings.

Finally, beamforming decoders were placed around the sphere as environmental "listeners". The decoders (referred to as MonoDecode) can be positioned utilizing a reduced spherical coordinate system. Additionally, the MonoDecoders can be given an azimuthal and elevational velocity, permitting automation of decoder movement. Examples of MonoDecoder control systems include controller mappings for the placement of decoder swarms, or stochastic processes which determine decoder trajectories. The decoder utilized in initial implementations utilized the typical sampling ambisonic decoding process defined previously in equation 2. As noted in the definition of SAD, the resulting signal is frequency-independent, and spatially bandlimited.

5 Creative Applications of Spherical Harmonic Projections in Timbre Mapping

5.1 Environmental Listener Metrics

Listener perception metrics have already been discussed in the first section of this paper, an intensive description of such metrics can be found in the works of Müller.¹⁴ Suppose a collection of audio samples containing a soprano singing excerpts from Ligeti’s *Mysteries of the Macabre* is placed all around the sound sphere. As these samples are played back, spectral measures such as Mel-frequency cepstrum coefficients (MFCCs)¹⁵ and the spectral kurtosis¹⁶ are taken. Now suppose that four independent MonoDecoders are traversing the sphere with their own respective trajectories. As they decode the spherical signal into a monophonic one, the same spectral measures are taken upon it. These measures are subsequently compared with the nearest soprano, and the absolute difference¹⁷ between the two is taken. This difference is then used to control a parameter to a spectral distortion processor, thus, as the decoders move further away from the nearest vocalist, their respective outputs become more and more distorted. Perception-based distance functions can be utilized to control a wide array of signal processors and control parameters, fueling an autonomous system which reacts to itself. Lastly, STO offers an intermediate space, analogous to the ambisonic b-format signal. In contrast to Wessel’s work, this space is unique as it can be transmitted to others, compressed, distorted, and modified to abstract its original makeup.

6 Commentary on Spherical Timbre Organization

6.1 Comparisons to Wessel’s Timbre Space Methods

Previously mentioned was Wessel’s method of the *space-master*. STO and Wessel’s work share an important design principle which is that the user must define where in space timbres or parameters must live. This decision was explicitly made in the creation of STO because it is influenced by acoustic ecology and soundscape creation. While Wessel’s work operates at the control level, STO can function using both audio and control data. STO offers a method of interpolation for data and mixing for audio sources, with the ability for users to define interpolation curves. This user specificity is not unique, however, as Wessel’s *space-master* provides the user a method for weighting each dimension.

14. M. Müller, *Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications*, Computer Science (Springer International Publishing, 2015), ISBN: 9783319219455, https://books.google.com/books?id=HCL_CgAAQBAJ.

15. Md. Sahidullah and Goutam Saha, “Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition,” *Speech Communication* 54, no. 4 (2012): 543–565, ISSN: 0167-6393, <https://doi.org/https://doi.org/10.1016/j.specom.2011.11.004>, <https://www.sciencedirect.com/science/article/pii/S0167639311001622>.

16. Wei Qiu, William Murphy, and Alice Suter, “Kurtosis: A New Tool for Noise Analysis,” *Acoustics Today* 16 (January 2020): 39–47, <https://doi.org/10.1121/AT.2020.16.4.39>.

17. The absolute difference between a and b being defined as $|a - b|$.

Wessel's work additionally has a solidified graphical interface, while STO's is still a work in progress.

6.2 Future Work

Further work on the theory and design of STO may include a well-defined graphical interface, pre-defined interpolation curves for audio and data sources, and possibilities for the mixture of audio and data sources within the same sphere. A concept currently being experimented with is *mixed-order* encoding and decoding. Using lower orders for ambisonic encoding produces a wider embedding upon the sphere; conversely, a high order encoding of sound upon the sphere produces a higher resolution placement. The inverse is found in decoding, where higher order decoding renders a narrower beam, while lower order decoding produces a wider beam. Through interpolation, this phenomenon can be adjusted smoothly, morphing from precise rays to wide ones.

STO is still in the early stages of its conception, as it currently only offers a structured *space*. While such an open-ended tool may blossom in the creative mind, a pre-defined set of tools or structures to be used within the space allows users to quickly apply and learn the technology without extensive practice or reading. The development of a concrete theory will facilitate experiments in spectral and temporal methods of decoding and encoding upon the sound sphere.

7 Conclusion

In conclusion, STO offers a new method of organizing and shaping timbre over time and space through a spherical interface. Rooted in ambisonic processing, it enables an already well-researched collection of effects and manipulations to be used. It additionally allows an intuitive connection to visuals which may convey to the audience any data which they are perceiving.

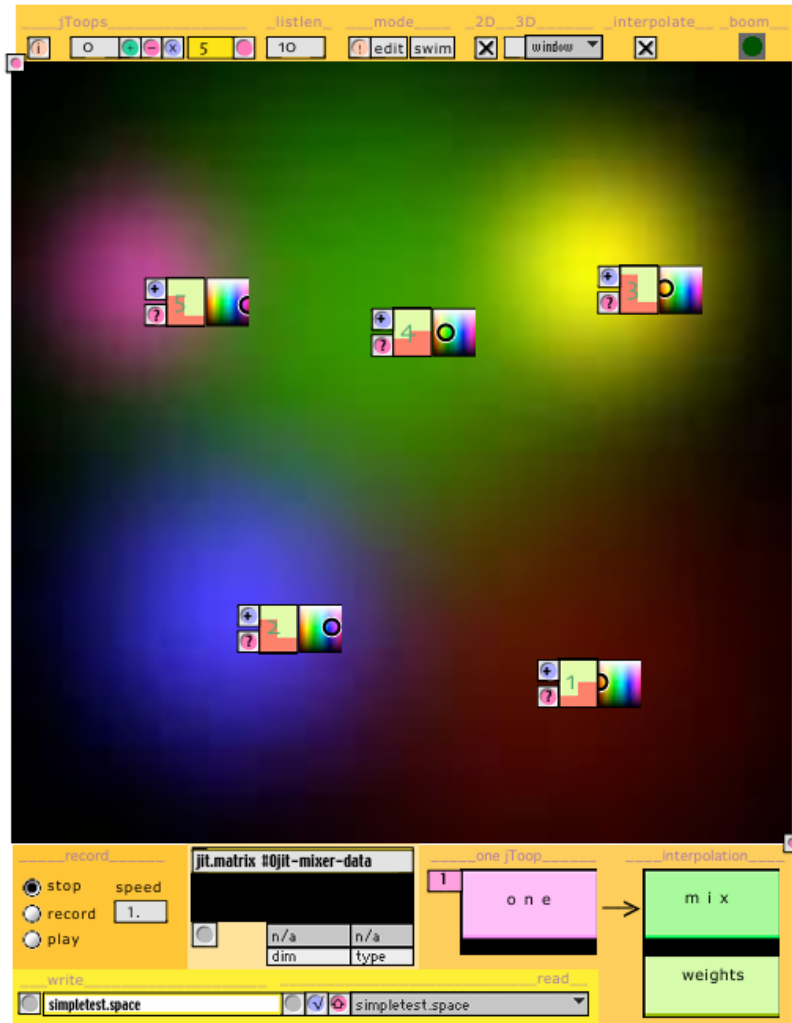


Figure 1: Wessel's *space-master* reproduced from his NIME paper on the software.¹⁹

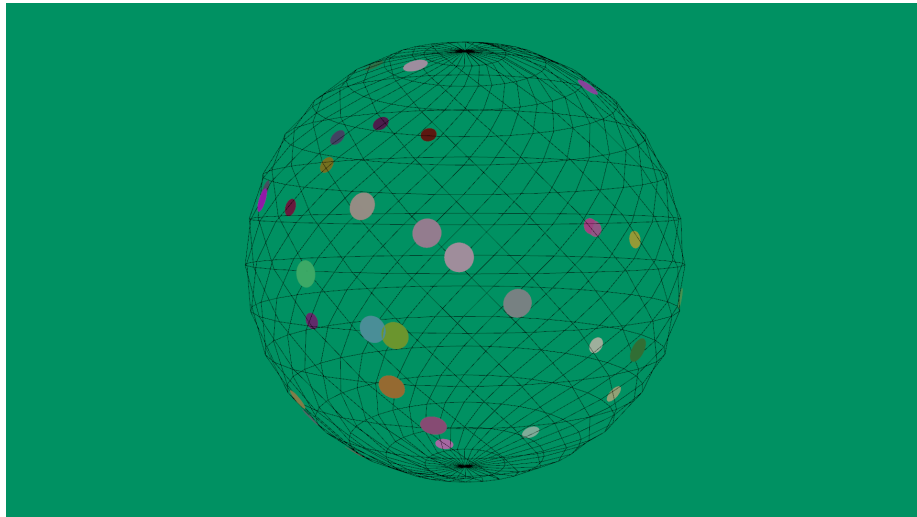


Figure 2: A visualization of all encoded sound files within the sound sphere.

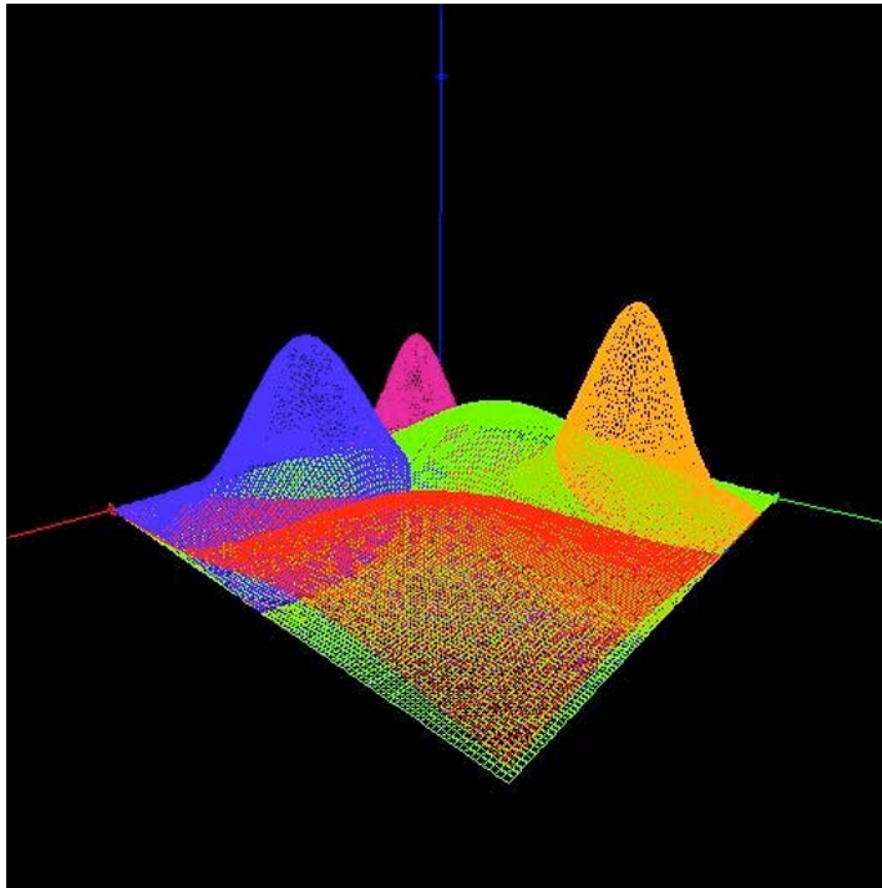


Figure 3: Wessel's visualization of timbre space placement from the NIME paper converging *space-master*.²¹

References

- Creasey, David, David Howard, and Andrew Tyrrell. “The Timbral Object - An Alternative Route to the Control of Timbre Space.” January 1996.
- Dworak, Paul E., and Philip C. Baczewski. “A Vector Field Model of Compositional Creativity.” January 1980.
- Frank., Matthias, and Frank Zotter. *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*. Creative Media Partners, LLC, 2020. ISBN: 9781013273193. <https://books.google.com/books?id=LET2zQEACAAJ>.
- Gregorio, Jeff, and Youngmoo E. Kim. “Evaluation of Timbre-Based Control of a Parametric Synthesizer.” <https://nime.pubpub.org/pub/adtb2zl5>, *International Conference on New Interfaces for Musical Expression*, April 2021.
- Grey, John M. “Multidimensional Perceptual Scaling of Musical Timbres.” *The Journal of the Acoustical Society of America* 61, no. 5 (May 1977): 1270–1277. ISSN: 0001-4966. <https://doi.org/10.1121/1.381428>. eprint: https://pubs.aip.org/asa/jasa/article-pdf/61/5/1270/11635616/1270__1__online.pdf. <https://doi.org/10.1121/1.381428>.
- Hoskinson, Reynald, and Dinesh Pai. “Manipulation and Resynthesis with Natural Grains.” January 2001.
- Kronlachner, Matthias. “Spatial Transformations for the Alteration of Ambisonic Recordings.” Master’s thesis, Institut für Elektronische Musik und Akustik, 2013.
- Lam, Joshua Ryan, and Charalampos Saitis. “The Timbre Explorer: A Synthesizer Interface for Educational Purposes and Perceptual Studies.” *International Conference on New Interfaces for Musical Expression*, April 2021.
- Mendelson, B. *Introduction to Topology: Third Edition*. Dover Books on Mathematics. Dover Publications, 2012. ISBN: 9780486135090. <https://books.google.com/books?id=FWFmoEUSwkC>.
- Momeni, Ali, and David Wessel. “Characterizing and Controlling Musical Material Intuitively with Geometric Models.” In *Proceedings of the International Conference on New Interfaces for Musical Expression*, 54–62, Montreal, Canada. Zenodo, June 2003. <https://doi.org/10.5281/zenodo.1176535>. <https://doi.org/10.5281/zenodo.1176535>.
- Müller, M. *Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications*. Computer Science. Springer International Publishing, 2015. ISBN: 9783319219455. https://books.google.com/books?id=HCI_CgAAQBAJ.
- Qiu, Wei, William Murphy, and Alice Suter. “Kurtosis: A New Tool for Noise Analysis.” *Acoustics Today* 16 (January 2020): 39–47. <https://doi.org/10.1121/AT.2020.16.4.39>.

- Sahidullah, Md., and Goutam Saha. "Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition." *Speech Communication* 54, no. 4 (2012): 543–565. ISSN: 0167-6393. <https://doi.org/https://doi.org/10.1016/j.specom.2011.11.004>. <https://www.sciencedirect.com/science/article/pii/S0167639311001622>.
- Samson, Séverine, Robert Zatorre, and James Ramsay. "Multidimensional Scaling of Synthetic Musical Timbre: Perception of Spectral and Temporal Characteristics." *Canadian journal of experimental psychology = Revue canadienne de psychologie expérimentale* 51 (January 1998): 307–15. <https://doi.org/10.1121/1.406009>.
- Terasawa, Hiroko, Malcolm Slaney, and Jonathan Berger. "A Statistical Model of Timbre Perception." In *ITRW on Statistical and Perceptual Audio Processing (SAPA 2006)*, 18–23. 2006.
- . "The Thirteen Colors of Timbre," 323–326. November 2005. ISBN: 0-7803-9154-3. <https://doi.org/10.1109/ASPAA.2005.1540234>.