

SAS and Multivariate Music: A GUI for Multivariate Analysis of Musical Instrumentation Digital Interface (MIDI) Data

**Lisa Cannon and Cecil Hallum
Department of Mathematics and Statistics
Sam Houston State University
Huntsville, Texas**

INTRODUCTION

Musical Instrument Digital Interface (MIDI) data provide an excellent and compact way to organize and use sound data; they comprise a relatively new way (since approximately 1982) to represent sound data, but their full potential has yet to be reached. Thus far, these files have found applications in areas where it would be advantageous to “hear” information in the data in addition to that which can be represented visually; clearly, studies of these applications are only in their infancy. In this paper, when these data are viewed in a linear algebraic framework and a form readable by SAS (Statistical Analysis System), extensive analyses and insights are possible.

The key structure of a MIDI file is simply that of a matrix whereby each row represents a clock beat in time of the frequencies played by a given collection of instruments (a maximum of 16 instruments comprise a MIDI file). Specifically, the first entry in the row

might be, for example, the frequency of play of an acoustic guitar, the second entry might be the frequency of play of a piano, etc. Consequently, all of Bach’s work is representable as a collection of matrices, M , where each matrix in M is a particular Bach piece. The same can be said about any musician, any genre of music, etc., whereby their music is in digital (i.e., MIDI) form.

This paper focuses on the linear algebraic structure of these MIDI data (i.e., digital music), and via SAS, establishes an initial statistical approach for characterizing and analyzing these files. An answer, for example, is provided to the question: “Can a Bach digital tune (i.e., a Bach MIDI) be transformed by matrix multiplication into a Beethoven MIDI tune”? Somewhat surprisingly, the answer is yes! In fact, this paper provides a complete characterization in regard to the algebraic characteristics that must exist between two MIDI matrices in

order for one to be transformable (i.e., be able to be mapped) to the other.

The emphasis of this paper, then, is on reporting the extent to which methods of multivariate statistical analysis and linear algebraic methods can be used to characterize the relationship between MIDI files.

OBJECTIVES

The SAS AF frame module was used to provide the opportunity to select a MIDI file, convert it from MIDI format to a format readable by SAS and to manipulate it in numerous ways to make statistical and algebraic conclusions. SAS code was also written to permit the file to be put back into a MIDI format to see how such manipulations affect the musical (or auditory) information by playing it back through a music synthesizer. For example, if one wanted to “listen” to the National anthem after transformation to principal component space, this later SAS program would permit such.

BACKGROUND

Oftentimes humans rely heavily on visual methods to communicate information (Mezrich, Frysinger, Slivjanovski, 1984); however, more effective use could possibly be made of

information that is expressible in the form of sound. A useful application of the sonification of data includes that of converting textured images to an auditory format (Martins, Rangayyan, Portela, Amero, Ruschioni, 1996). This allows for clarification of a visually crowded image, where too many visual variables are present, inhibiting the brain from sorting the information appropriately. Stoering (1996) has created a prototype of a helmet instrument containing a camera that converts visual information into sound data, that to an untrained ear would sound like static. A seeing impaired individual could learn to differentiate between the many frequencies and intensities of static to identify their surroundings without seeing them. Physicists and engineers have recently developed yet even another application of sound, a thermo-acoustic device that will convert sound waves into mechanical energy that is easily converted into a temperature differential (Krupnick, 2000).

Taking advantage of the relationship between music and mathematics, musical pieces can be created not only by the expert skill of a

highly trained composer, but also using known relationships in mathematics (Bulmer, 2000). Many toolkits have been created to allow the user to convert sound data into mathematical information for various purposes (Lodha, Beahan, Heppe, Joseph, Zane-Ulman, 1997). The Musical Sonification Environment (MUSE) converts visual information into sound information. Various visual variables are converted to different auditory variables including timbre, rhythm, volume, pitch, tempo and harmony. This particular toolkit is intended to assist scientific research in which visual information is cluttered resulting in uncertainty.

The classification of musical pieces into genres proposes a problem easily solved from the perspective of musical data (Pachet, Westermann, Laigre, 2001). Musical data mining using co-occurrence and correlation analysis can assist in this classification application as well.

The method of multidimensional scaling analysis was used to identify the physical description of musical instrument sounds in order to assign each instrument with a unique “tone signature” (Kaminskyj, 1999). By using

principal component analysis, the most significant attributes of each sound were determined.

Using sound to represent multivariate analytical data can prove to be more appropriate than visual data to assist an analyst in pattern recognition (Yeung, 1980). Despite the ability of statistical methods to objectively handle multi-dimensional data, human conducted pattern recognition is sometimes necessary. Listening to the multidimensional information before running statistical tests can reduce computational costs by allowing the study to focus only on parameters that display an auditory pattern. Pitch, loudness, damping, direction, duration/repetition, and rest are musical sound characteristics that the human ear, with very little training can learn to differentiate.

The automated analysis of musical sources, called Musical Pattern Discovery, can be used to mimic the human perception of music in order to determine pattern classes to identify the chorus of a piece. Lartillot (2003) proposed a method that mimics a listener’s cognitive and sub-cognitive inferences when hearing a piece.

Chai and Vercoe (2003) found a method to statistically analyze an acoustical piece of music in order to index and thumbnail the piece. To determine the musical structure of a piece, this procedure sought similar segments in a musical piece, clustered the similar segments, and interpreted the musical structure in terms of the clusters.

Das, Howard, and Smith (1999) of the University of New York have conducted a limited study on multivariate MIDI data applied to musical kinematics. They applied multivariate methods to investigate the motion of the music, such as velocity and tempo changes. They found that different motion types have different velocity profiles that can be expressed mathematically and analyzed statistically.

APPROACH

As a part of this effort, a SAS program was developed that converts MIDI files into SAS readable form. A user is permitted to select a MIDI tune from a list using the SAS AF module.

For this paper, several tunes by David Bowie found on the internet at <http://www.bowiewonderworld.com/baby.htm> are compared. To select a MIDI

tune, first the user must select a category (using the AF screen in Figure 1). Here D_Bowie has been selected, which automatically changed the entries in the “MIDI FILES” list box to only tunes by David Bowie. *Space Oddity*, a David Bowie tune, has been selected from this menu. Clicking on “APPLY” transforms the MIDI file into a SAS data file. At this point the user may choose to view the MIDI file by clicking on “VIEW MIDI”, which brings up the AF frame screen shown in Figure 2.

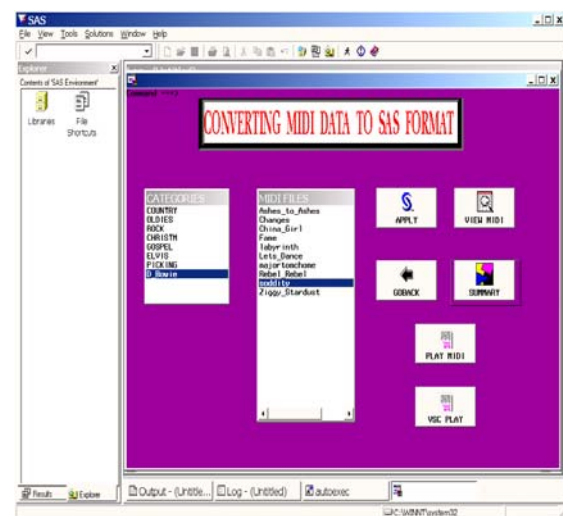


Figure 1: Converting MIDI Data to SAS Format Frame.

The user may also choose to listen to the MIDI file by selecting “Play MIDI,” which automatically pulls up the Windows Media Player, or a music synthesizer “VSC PLAY,” which pulls

up Roland's Virtual Sound Card synthesizer. In order to complete the conversion to a matrix, click on the "GOBACK" button to return to the SAS editor. Once back in the SAS editor, a particular SAS program can be run to strip the information from the MIDI files and reorganize it into a matrix format such that the information for each instrument is restricted to an assigned column with all of the clock beats aligned to match the timing of all other instruments present in a tune. The code for this program will be handed out at conference time; this program goes through each element of the MIDI file organizing the information carefully into columns of instruments and rows corresponding to clock beats, creating a

matrix that can be manipulated by PROC IML in SAS. This PROC IML program is a basic program that creates the actual MIDI matrix by converting the organized data from MIDI pitch code (0 to 127) into standard sound frequencies and then inputs the information into a matrix which can easily be manipulated. The following SAS source code is included in the transformation program, which permanently stores the matrix M created by converting the song *Space Oddity* into a matrix called M_0 .

$M0 = M$; restore = sasuser.imlstor; store $M0$;

This procedure is then repeated for the rest of the David Bowie songs available. Each new matrix is saved under a different name as shown in Table 1.

Table 1: David Bowie MIDI Tunes

Song Title	MIDI Title	Stored As
Space Oddity	Soddity	M0
Ashes to Ashes	Ashes_toAshes	M1
Changes	Changes	M2
China Girl	China_Girl	M3
Fame	Fame	M4
Labyrinth	Labyrinth	M5
Let's Dance	Lets_Dance	M6
Rebel Rebel	Rebel_Rebel	M8

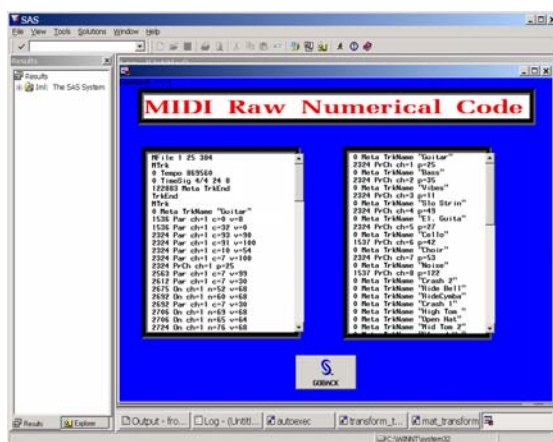


Figure 2: MIDI Raw Numerical Code Frame resulting from clicking on "View MIDI."

Once the matrix transformation program has been run, another SAS program was

written to transform back to the MIDI format for listening purposes. After converting the matrix to MIDI, the Windows Media player automatically opens, allowing the user to listen to the results. Again after the matrix is manipulated in various ways, additional SAS Screen Control Language (SCL) code allows the user to hear how the manipulation changes musical and auditory properties of the manipulated song.

TECHNICAL ASPECTS

Now, assume that one has two MIDI matrices created using the IML program; M_0 is a $m \times n$ matrix having m clock beats and n instruments, and M_i is a $p \times n$ matrix having p clock beats and n instruments. Particularly, note also that MIDI tunes have a special structure whereby each component in every column is restricted to one of only 128 MIDI note values that correspond to 128 frequencies. These 128 frequencies cover approximately a ten octave range. However, most tunes stay within a two or three octave range, which means each of the column values are further restricted (in the three octave case) to no more than 36 distinct values. Given that

a MIDI tune is in a particular key, the values are limited even more to being only within the diatonic scale of three octaves (21 values). Often, not all of these possible notes are utilized in the tune. Thus, the possible values present in each column are relatively small. A critical part of the approach utilized herein relies on the linear matrix model relationship between MIDI tunes (i.e., M_0 and M_i) expressed by $M_0 = TM_i + E$, where the expected value of the random error matrix E is assumed to be zero. The key question is: “What are the characteristics of a transformation T that results in changing one MIDI tune into another?”. The distance utilized herein between two matrices (here M_0 and TM_i) is the Frobenius norm of M_0 and TM_i , defined (using the trace operator “Tr”) as $\sqrt{\text{Tr}((M_0 - TM_i)'(M_0 - TM_i))}$. In this expression, the prime is the matrix transpose operator. Notice that $M_0 - TM_i$ is simply the error associated with the transformation of M_i into M_0 . In order to find an appropriate transformation, T , of M_i , the transformation, T , is selected that

minimizes the square of the Frobenius norm,

$$Tr(E'E) = Tr(M_0 - TM_i)'(M_0 - TM_i),$$

which essentially minimizes the error (Schott, 1997). This is achieved by obtaining the matrix differential with respect to T of the square of the Frobenius norm and equating this to zero. This results in the matrix equation $TM_i M_i' = M_0 M_i'$ which has every solution given by the expression

$$T = M_0 M_i' (M_i M_i')^+ + Z' (I - M_i M_i' (M_i M_i')^+)$$

where the superscript “+” denotes the matrix pseudoinverse (see Schott (1996)). Multiplying by M_i on the right in the above expression results in $TM_i = M_0 M_i^+ M_i + Z' (M_i - M_i) = M_0 M_i^+ M_i$. Note that if M_i is of full column rank,

$$M_i^+ = (M_i' M_i)^{-1} M_i', \text{ which implies that}$$

$$TM_i = M_0, \text{ or } E = 0. \text{ Consequently,}$$

we have Result 1.

Result 1: Given the linear relationship between two MIDI tunes $M_0 = TM_i + E$, the transformation T which minimizes $Tr(E'E)$ is given by

$$T = M_0 M_i' (M_i M_i')^+ + Z' (I - M_i M_i' (M_i M_i')^+)$$

where Z is an arbitrary matrix conformable for multiplication and the

estimator of E is $\hat{E}$ where

$$\hat{E} = M_0 - TM_i = M_0 (I - M_i^+ M_i) \text{ and}$$

$TM_i = M_0 M_i^+ M_i$. Consequently, if the column space of the MIDI tune M_0 is contained in the column space of M_i then $\hat{E} = 0$.

In order to assure that the MIDI matrices M_0 and M_i are conformable for matrix multiplication, additional SAS code given below extends the dimensions of M_i so that the number of columns in M_i is the same as the number of rows in M_0 filling in the remaining entries with zeros. Effectively this is equivalent to adding “muted” instruments that do not contribute any new musical information: `newm1 = shape[nrow(M1), ncol(M0), 0];` while this does not add new information, it ensures that M_0 and TM_i are conformable for matrix addition. If M_0 has fewer columns than M_i , a subset of columns of M_i can be removed, which musically amounts to selecting a subset of the instruments.

The principal components of the MIDI matrix permit ranking the sources of variation attributable to the

instruments that compose the matrix columns. Running a principal component analysis of the covariance matrix of a MIDI tune in the SAS yields all the eigenvalues and associated eigenvectors for the tune. SAS ranks the eigenvalues, indicating the most influential components and how much of the variability in the tune is explained by that particular component. The eigenvectors can be used to determine how influential each instrument is by investigating the eigenvector components. The instrument corresponding to the eigenvector component having the largest absolute value (associated with the largest eigenvalue) is the most influential. In all cases investigated to-date, the first principal component selects the melody instrument.

Even though the first principal component sorts out the most influential instrument, a tune is not necessarily completely defined by its melody. The remaining instruments typically contribute more useful and interesting information that is characteristic and unique to the piece of music. We observed that having 90.00% of a tune explained by the principal components

(as measured by the cumulative proportion of eigenvalues) suffices to identify which instruments uniquely characterize a MIDI tune.

One might now ask: “What effect does multiplying M_i on the right by an arbitrary nonsingular matrix C have on the relationship between two matrices? Consider the transformation of the MIDI matrix M_i to $M_i C$. Then the transformation matrix becomes

$$T = M_0 (C' M_i' M_i C)^+ C' M_i',$$

which yields

$$T M_i C = M_0 (C' M_i' M_i C)^+ C' M_i' M_i C$$

In this case, the transformation is not invariant to this kind of alteration unless $M_i C$ is of full column rank. Thus the question arises as to whether there is some condition on C whereby multiplication on the right results in invariance (i.e., $T M_i C = T M_i$). Note that if C is an orthogonal matrix, then $C' = C^{-1}$. Then $T M_i C$ reduces to

$T M_i C = M_0 C^{-1} (M_i' M_i)^+ (C')^{-1} C' M_i' M_i C$ after distributing the pseudoinverse across the parentheses, which becomes

$$T M_i C = M_0 C^{-1} (M_i' M_i)^+ M_i' M_i C.$$

For this to reduce further, consider that $(M_i' M_i)^+ = (M_i' M_i)^{-1}$, which is possible

only if M_i is of full column rank. Then $TM_iC = M_0C^{-1}(M_i'M_i)^{-1}M_i'M_iC = M_0C^{-1}C = M_0$. This implies that if M_i is of full column rank and C is orthogonal, then the transformation of M_iC into the MIDI matrix M_0 results with no error. Note that in order to satisfy the condition that $C' = C^{-1}$, it will suffice for C to be a permutation matrix. Also note that the transformation of M_i into M_0 will result with no error if the MIDI matrix M_i is of full column rank. This yields Result 2, a special case of Result 1.

Result 2: The transformation whereby matrix M_i is multiplied by a matrix C on the right, $M_0 = TM_iC + E$, does not affect the relationship between M_0 and M_i provided C is an orthogonal matrix. Further explanation of this will be given at conference time.

CONCLUSIONS

It is hoped that this paper serves to establish an initial base in the linear algebraic and multivariate structure of MIDI files and also prompt continued research to expand the analysis and manipulation of sound data. The

applications of such will, expectedly, prove useful in extensively analyzing information of sound data. Results of the transformations on the David Bowie tunes and others will be discussed in considerable detail at conference time. In particular, actual MIDI tunes will be discussed whereby the conditions of Result 1 hold true and whereby they do not hold. PC Speakers will be set up at conference time to permit listening to transformed files; e-mail addresses will be provided whereby additional examples are included. All AF screens and their associated screen control language (SCL) will be handed out at conference time.

REFERENCES

- Bulmer, M. (2000), "Music from Fractal Noise," Proceedings of the Mathematics 2000, Festival, Melbourne.
- Chai, W., and Vercoe, B. (2003), "Structural analysis of Musical Signals for Indexing and Thumbnailing," Proceedings of the International Multimedia Conference, Berkeley, CA.
- Chew, E., and Chen, Y. (2003), "Mapping MIDI to the Spiral Array: Disambiguating Pitch Spellings," Proceedings of the International Multimedia Conference, Berkeley, CA.
- Das, M., Howard, D., and Smith, S. (1999), "Motion Curves in Music: the Statistical Analysis of MIDI Data,"

- Proceedings of the 25th Euromicro Conference, Milan.
- Dubnov, S., Assayag, G., Lartillot, O., and Bejerano, G. (2003), "Using Machine-Learning Methods for Musical Style Modeling," *Computer*, 36, 73-80.
- Felippa, C. (1986), "Introduction to Finite Element Methods," available at <http://caswww.colorado.edu/courses.d/I/FEM.d/Home.html>.
- Foote, J., Cooper, M., and Nam, U. (2002), "Audio Retrieval by Rhythmic Similarity," Proceedings of the International Conference of Music Information Retrieval, Paris.
- Johnson, R., and Wichern, D. (2002), *Applied Multivariate Statistical Analysis* (5th ed.), Upper Saddle River, NJ: Prentice Hall.
- Kaminsky, I. (1999), "Multidimensional Scaling Analysis of Musical Instrument Sounds' Spectra," Proceedings of the Australian Computer Music Conference, Wellington.
- Kinder, P. (2002), "Bowie Wonderworld.Com," available at <http://www.bowiewonderworld.com/baby.htm>.
- Krupnick, S. (2000), "Smaller Lasers Demand New Approaches to Chiller Design," *Laser World Focus*, 36, Issue 11, available at <http://lfw.pennet.com/home.cfm>.
- Krygier, J. (1994), "Sound and Geographic Visualization," *Visualization in Modern Cartography*, New York: Pergamon, 149-166.
- Lartillot, O. (2003), "Perception-Based Musical Pattern Discovery," Proceedings of the International Computer Music Conference, Singapore.
- Lodha, S., Beachan, J., Heppe, T., Joseph, A., and Zane-Ulman, B. (1997), "MUSE: A Musical Data Sonification Toolkit," Proceedings of the International Conference on Auditory Display, Palo Alto, CA.
- Martins, A., Portela, L., Rangayyan, R., Amaro, E., and Ruschioni, R. (1996), "Auditory Display and Sonification of Textured Image," Proceedings of the International Conference on Auditory Display, Palo Alto, CA.
- Meudic, B. (2003), "Musical Similarity in a Polyphonic Context: a Model Outside Time," Proceedings of the XIV Colloquium on Musical Informatics, Firenze, Italy.
- Mezrich, J., Frysinger, S., and Slivjanovski, R. (1984), "Dynamic Representation of Multivariate Time Series Data," *Journal of the American Statistical Association*, 79, 34-40.
- Moon, T., and Wynn, S. (1999), *Mathematical Methods and Algorithms for Signal Processing*. Upper Saddle River, NJ: Prentice Hall.
- Myers, R., and Milton, J. (1998), *A First Course in the Theory of Statistical Models* (2nd ed.), New York: McGraw-Hill.
- Pachet, F., Westermann, G., and Laigre, D. (2001), "Musical Data Mining for Electronic Music Distribution," Proceedings of the First International

Conference on WEB Delivering of Music, Florence.

Pauletto, S., and Hunt, A. (2004), "Interactive Sonification in Two Domains: Helicopter Flight and Physiotherapy Movement Analysis," Proceedings of the International Workshop on Interactive Sonification, Bielefeld, Germany.

Pickens, J. (2000), "A Survey of Feature Selection Techniques for Music Information Retrieval," Proceedings of the International Symposium on Music Information Retrieval, Plymouth, MA.

Schott, J. (1997), *Matrix Analysis for Statistics*, New York: John Wiley and Sons.

Serra, X. (1997), "Musical Sounds Modeling with Sinusoids Plus Noise," *Musical Signal Processing*, eds. C. Roads, S. Pope, A. Piccialli, and G. De Poli, Barcelona: Swets and Zeitlinger Publishers.

Stoering, P. (1996), "The VOICE Learning Edition," available at <http://www.seeingwithsound.com/winvoice.htm>.

Tzanetakis, G., Ermolinskyi, A., and Cook, P. (2002), "Pitch Histograms in Audio and Symbolic Music Information Retrieval," Proceedings of the International Computer Music Conference, Gothenburg, Sweden.

Yeung, E. (1980), "Pattern Recognition by Audio Representation of Multivariate Analytical Data," *Analytic Chemistry*, 52,1120-1123.