

Independent component analysis for automatic note extraction from musical trills^{a)}

Judith C. Brown^{b)}

Physics Department, Wellesley College, Wellesley, Massachusetts 02181 and Media Lab, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Paris Smaragdis^{c)}

Mitsubishi Electric Research Lab, Cambridge, Massachusetts 02139

(Received 19 June 2003; accepted for publication 9 February 2004)

The method of principal component analysis, which is based on second-order statistics (or linear independence), has long been used for redundancy reduction of audio data. The more recent technique of independent component analysis, enforcing much stricter statistical criteria based on higher-order statistical independence, is introduced and shown to be far superior in separating independent musical sources. This theory has been applied to piano trills and a database of trill rates was assembled from experiments with a computer-driven piano, recordings of a professional pianist, and commercially available compact disks. The method of independent component analysis has thus been shown to be an outstanding, effective means of automatically extracting interesting musical information from a sea of redundant data. © 2004 Acoustical Society of America.

[DOI: 10.1121/1.1698774]

PACS numbers: 43.75.St, 43.60.Cg, 43.75.Mn [SEM]

Pages: 2295–2306

I. INTRODUCTION

As with many fields today the processing of digitized musical information is inundated with huge masses of data, much of which is redundant. One attempt to reduce this deluge of data in the musical domain was attempted a decade ago with principal component analysis or PCA (Stapleton and Bass, 1988; Sandell and Martens, 1995). Earlier Kramer and Mathews (1956) had written an excellent introduction to data reduction in audio.

Since then the field of information processing has made a stride forward with new algorithms, one of particular interest to audio being independent component analysis [ICA; see Hyvarinen (1999) for an excellent introduction]. This method has been used with success in what is called blind source separation (Torkkola, 1999) and is a solution under certain restrictions to the computational statement of the age old cocktail party effect, addressing the question of whether a machine can emulate a human in picking out a single voice in the presence of other sources. The solution to this problem is considered by many to be the holy grail of audio signal processing. A restriction in the mainstream use of ICA has been that the number of microphones must be equal to or greater than the number of sources.

Recent reports (Casey and Westner, 2000; Smaragdis, 2001; Brown and Smaragdis, 2002) have indicated that, if a signal is preprocessed into frames of magnitude spectral features, then independent component analysis can be applied without the constraint of multiple microphones to extract the features carrying maximum information. We will develop

this method further for the analysis of trills with a twofold purpose:

- (1) Automatic redundancy reduction—We will show that ICA can be used to obtain musical information quickly, easily, and accurately from data recorded with a single microphone.
- (2) Creation of database—From the calculations with ICA, we will assemble a database of information on a large number of trills obtained from a variety of sources to draw conclusions about trill rates.

II. BACKGROUND

A. Statistics background

Most of the sensory information we receive is highly redundant, and the goal of acoustical signal processing is often to expose the fundamental information and disregard redundant data. Since this is a common problem in data processing, statistical methods have been devised to deal with it. The following sections describe two of the most powerful techniques applied to spectral audio data.

1. Principal component analysis

A number of data reduction techniques are based on finding eigenfunctions for the second-order statistics of the data (Therrien, 1989). These techniques attempt to approximate a given data set using the superposition of a set of linearly independent functions, called basis functions, in a manner similar to the approximation of a sound by the superposition of sinusoids. Using a number of basis functions that equals the dimensionality of the original data set gives a perfect reconstruction. More often, the use of a reduced set of these functions results in efficient data encoding or a more useful interpretation of the data. The most prominent of these

^{a)}Portions of these results were first presented at the 143rd ASA Meeting in Pittsburgh, PA (Brown and Smaragdis, 2002).

^{b)}Electronic mail: brown@media.mit.edu

^{c)}Electronic mail: paris@merl.com

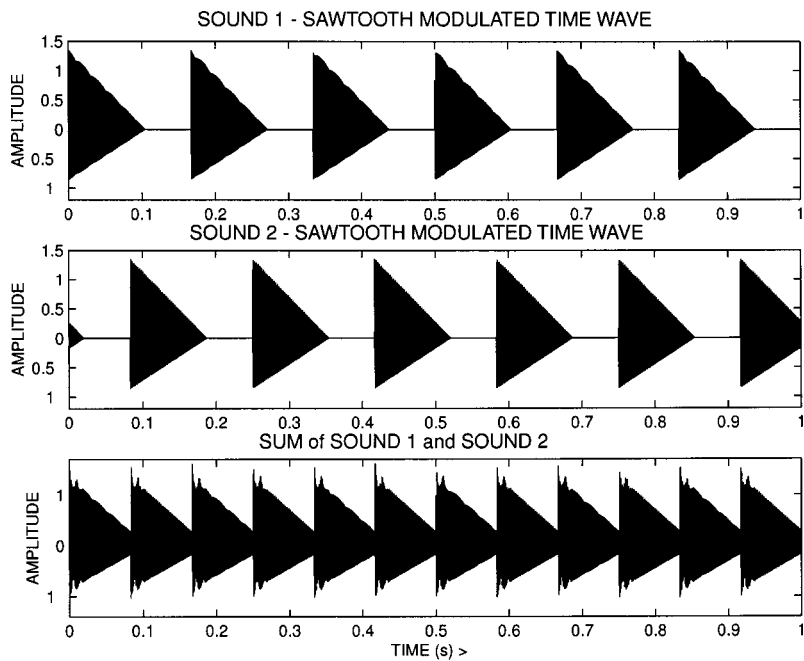


FIG. 1. Synthetic signal simulating a trill and consisting of the sum of two complex sounds, each containing three harmonics and modulated by a low-frequency sawtooth. The upper two graphs are the individual complex sounds, and the bottom graph is the sum.

approaches is called principal component analysis in the statistics literature, also referred to as the Karhunen–Loeve transform in the signal processing literature.

More formally, given a set of data vectors of dimension N , the method of principal component analysis can be used to find new a set of N' ($N' \leq N$) basis functions which are uncorrelated (second-order independence) and can be used to reconstruct the input. These are optimal in the sense that no other set of N' vectors gives a better least mean squares fit. The new basis functions can be sorted by magnitude of their variance, which is a measure of their importance in describing the data set. Optionally we can ignore the least important bases, and the dimensionality of the data set can be reduced with fine detail eliminated.

As an example applicable to our later sections, we consider the matrix of values calculated for the magnitude of the constant-Q transform (Fourier transform with log-frequency spacing) of a temporal waveform broken up into N shorter time segments. The calculation was carried out by the method of Brown (1991; Brown and Puckette, 1992) with a Q of 17 corresponding to the frequencies of musical notes. The time wave is the sum of two synthetic sounds with fundamental frequencies corresponding to musical notes C_6 and D_6 and each containing harmonics two and three. These sounds are amplitude modulated by a low-frequency sawtooth simulating alternating notes as found in trills (Fig. 1).

Figure 2 is a plot of the constant-Q coefficients calculated for the input time wave of Fig. 1. Each column repre-

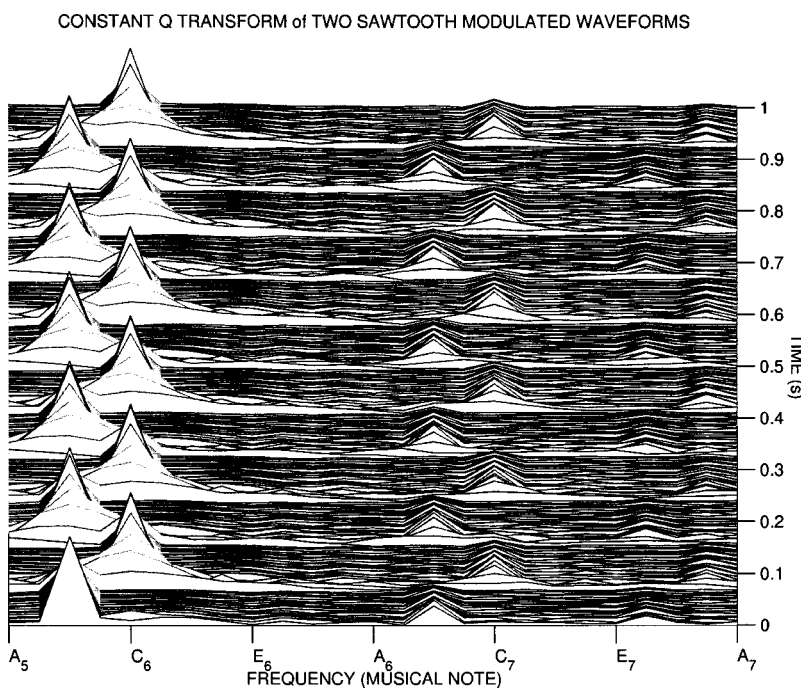


FIG. 2. Magnitude (arbitrary units) of the constant-Q transform against frequency and time in seconds (waterfall plot) for the complex sound of Fig. 1. Frequencies are indicated on the horizontal axis by musical notes.

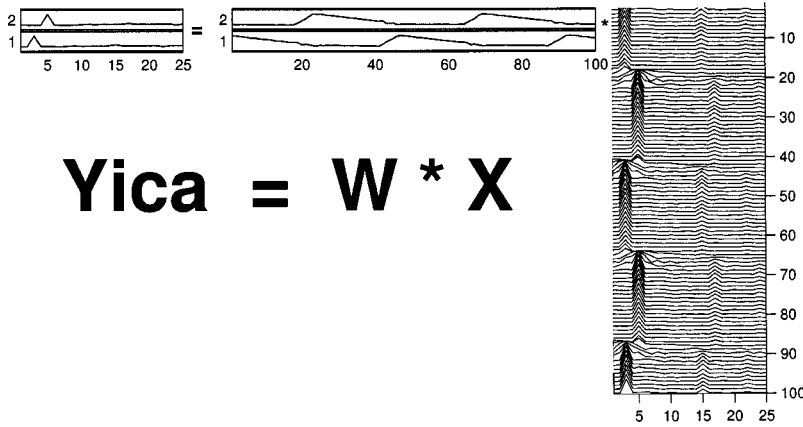


FIG. 3. Graphical example of the matrix multiplication $Y = W * X$ for the first third of the data matrix of Fig. 2, keeping the two most important independent components. The independent components Y for this orientation of the data matrix are the frequency bases. Note that the two basic shapes of the rows of X have been extracted. The transformation matrix W displays the temporal behavior of the two independent components (same shape as columns three and five) and are referred to as time bases.

sents the values of one spectral coefficient at N times, and each row consists of M frequency samples of a single variate. Viewed as a whole, the columns are components of a random vector, and each column is a sample of that vector at a different frequency. These data are highly redundant with one basic shape for the spectra of the two notes present differing only in their horizontal positions.

It is more common to consider the transpose of this matrix, which gives samples in time for the rows, but better results were obtained as described. This is because the frequency-dependent rows (or samples) are better separated and hence less correlated for the covariance calculation.

Subtracting the average of each row from the elements of that row, and defining a typical element of the covariance matrix C (Therrien, 1989) as the expectation value, we have

$$C_{ij} = \langle X_i \cdot X_j \rangle, \quad (1)$$

where the average is taken over all samples. See Appendix A for an example of this implementation.

For a finite data set where all samples are available as rows in a matrix X , the covariance matrix can be computed by

$$C = \langle X \cdot X^T \rangle, \quad (2)$$

where X^T is the transpose of X . This matrix can be diagonalized by finding the unitary transformation U such that

$$U^T \cdot C \cdot U = D, \quad (3)$$

where D is diagonal. This is done by solving for the eigenvalues of C with the result

$$U^T \cdot (X \cdot X^T) \cdot U = D. \quad (4)$$

From Eq. (4) using the associative property of matrices and the transpose of a product

$$(U^T \cdot X) \cdot (X^T \cdot U) = (U^T \cdot X) \cdot (U^T \cdot X)^T. \quad (5)$$

Defining a new matrix in Eq. (5),

$$Y_{pca} = U^T \cdot X. \quad (6)$$

Y_{pca} is the matrix of principal components (called scores in the statistics literature) and has a diagonal covariance matrix with elements equal to the variances of its components. U^T is called the weights matrix in the statistics literature. Both Y_{pca} and the transformation matrix U^T can be ordered by magnitude of the variance. The dimensionality can thus be reduced

by taking the k rows of each of these matrices corresponding to the largest variances. See Fig. 3 for an example of matrix multiplication keeping two components.

With this orientation of the data matrix X , the rows of Y_{pca} will be spectra corresponding to the rows with the largest variances and will be referred to as frequency bases. See the frequency dependence for the two complex sounds in Fig. 2. The rows of U^T , the unitary transformation matrix, will show the time dependence for the k most important rows and be referred to as time bases.

Since the covariance matrix of Y_{pca} is diagonal, off-diagonal elements

$$D_{ij} = \langle Y_i \cdot Y_j \rangle \quad (7)$$

are zero showing that the components of Y are orthogonal or linearly independent. From a statistical point of view they are decorrelated showing

$$E[Y_i Y_j] = E[Y_i] \cdot E[Y_j] = 0. \quad (8)$$

This form of independence does not, however, mean that the two components are completely uncoupled and that they are statistically independent. For true statistical independence the joint probability density must factor into the marginal densities

$$p(Y_i, Y_j) = p(Y_i) \cdot p(Y_j), \quad (9)$$

and for this factorization to hold another method is needed.

2. Independent component analysis

The goal of independent component analysis is to find a linear transform

$$Y = W \cdot X \quad (10)$$

such that the variates of Y are maximally independent. Stated otherwise, this transform should make the equation

$$p(Y_1, \dots, Y_M) = \prod_{i=1}^M p(Y_i) \quad (11)$$

“as true as possible”. It is much more difficult to find the desired transformation W than the corresponding unitary transformation for PCA. One approach has been to minimize the relative entropy or Kuhlback–Liebler (KL) divergence (Deco and Obradovic, 1996). This is a quantity defined in information theory to give a measure of the difference in two

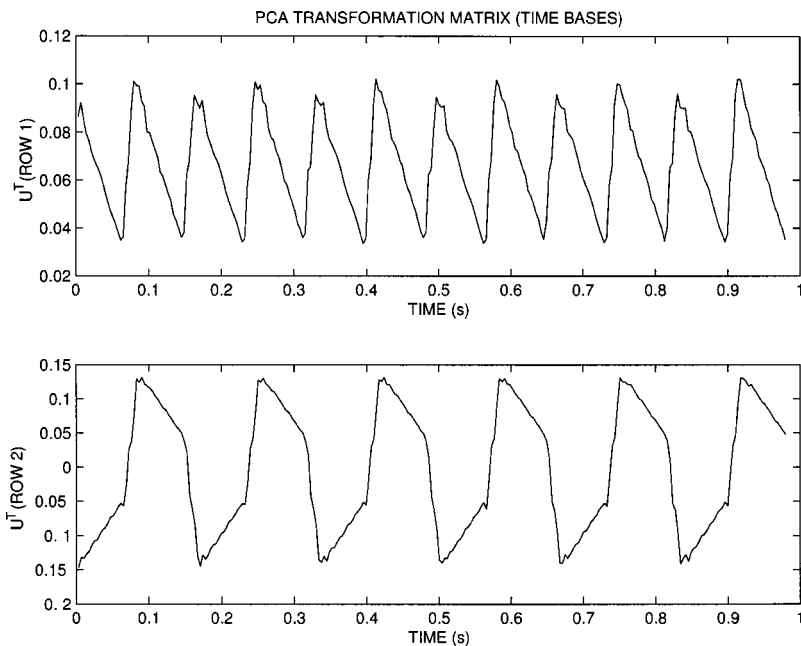


FIG. 4. PCA transformation matrix—the two most important rows of the unitary transformation matrix U^T of Eq. (6) for the complex sound of Fig. 1 with the constant-Q transform of Fig. 2 as data matrix X .

probability densities and has been used extensively for pattern classification.

The KL divergence is defined for two probability densities $p(x)$ and $q(x)$

$$K(p||q) = \int p(x) \log\left(\frac{p(x)}{q(x)}\right) dx, \quad (12)$$

where the integral is taken over all x . The KL divergence can be easily adapted as a measure of the difference in the joint probability and the marginal densities in Eq. (11). In this context it is called the mutual information (Deco and Obradovic, 1996) $I(Y_1; \dots; Y_M)$ and is a measure of the statistical independence of the variates whose densities appear on the right side of Eq. (11). That is, it tells us to what degree the Y_i are statistically independent:

$$I(Y_1; \dots; Y_M) = K\left(p(Y_1, \dots, Y_M) \parallel \prod_{i=1}^M p(Y_i)\right). \quad (13)$$

Several algorithms for ICA solutions have used procedures which have the effect of minimizing the mutual information including those of Amari (1996) and Bell and Sejnowski (1995). These are called infomax and in general seek a transformation matrix \mathbf{W} in Eq. (10) in an iterative calculation.

An alternative approach, which is conceptually close to PCA, is to extend the second-order independence of PCA to higher orders using a cumulant-based method. This is the approach taken by Cardoso (1990; Cardoso and Souloumiac, 1996) in diagonalizing the quadricovariance tensor. Instead of the terms C_{ij} of the covariance matrix, he considers all products up to fourth order such as

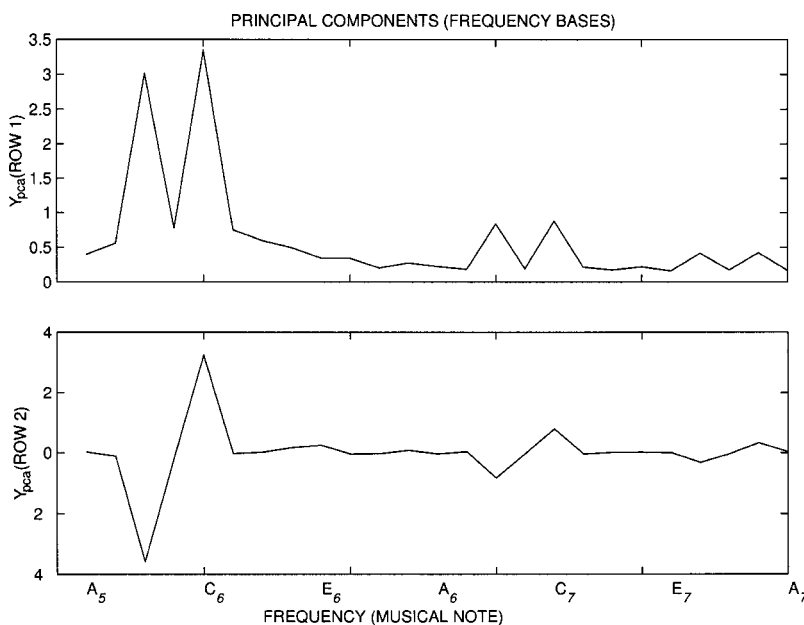


FIG. 5. The two most important principal components \mathbf{Y}_{pca} from the transformation equation (6) for the complex sound of Fig. 1 with the constant-Q transform of Fig. 2 as data matrix X . Frequencies are indicated by musical notes on the horizontal axis. Note that the two basic shapes of Fig. 2 are mixed by the transformation.

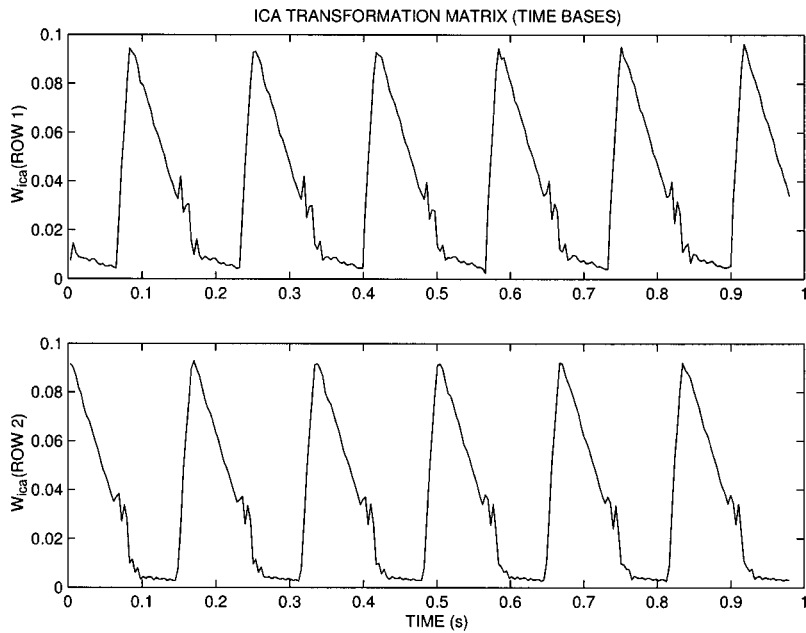


FIG. 6. ICA Transformation matrix—the two most important rows from the transformation matrix \mathbf{W} of Eq. (10) for the complex sound graphed in Fig. 1 with the constant-Q transform of Fig. 2 as data matrix X . This is also the first third of the matrix W in Fig. 3.

$$C_{ijkl} = \langle X_i \cdot X_j \cdot X_k \cdot X_l \rangle. \quad (14)$$

The diagonalization of this tensor ensures that no two dimensions of the data will have a statistical dependence up to and including the fourth order. This is a generalization of the diagonalization of the covariance matrix as done with PCA, where dependencies are eliminated up to second order. By extending the notion of the covariance matrix and forming the quadricovariance tensor (a fourth-order version of covariance), we effectively set a more stringent definition of statistical independence.

This concept can also be extended to an arbitrary order of independence by forming and diagonalizing even more complex structures. In this case the complexity of the process unfolds exponentially and can present computational issues. Fourth-order independence is a good compromise, ex-

hibiting a manageable computational burden with good results.

B. Trill background

Trills were chosen for this study because they are extremely difficult to analyze. The note rate is very rapid, and when pedaled there are two temporally overlapping notes present. There is an advantage, however, in that they do not have simultaneous onsets.

The execution of trills has been studied by a number of groups interested either in performance on musical instruments or in perception limits of detection of two pure tones. The latter measurements are best summarized by Shonle and Horan (1976) who varied the frequency difference of two sinusoids with a modulation rate (frequency of a trill pair) of

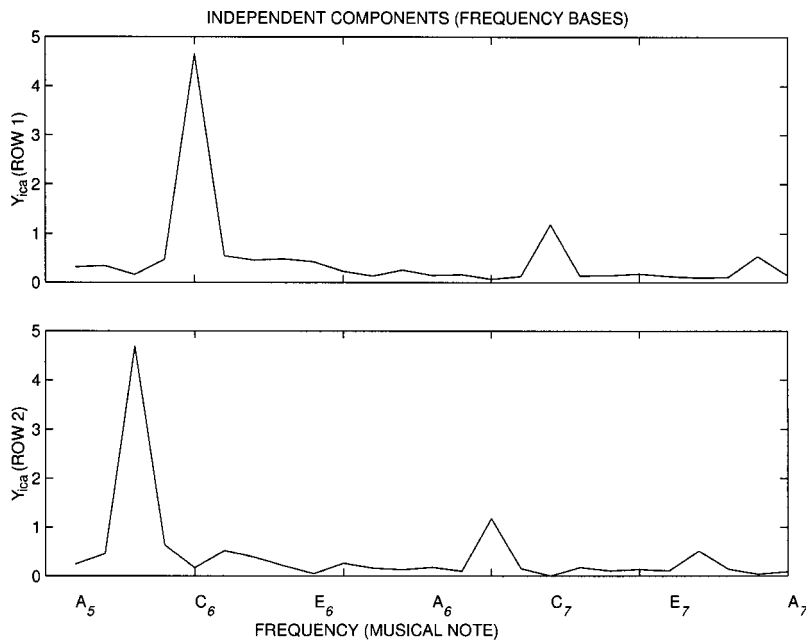


FIG. 7. Independent components—the two most important rows of the matrix \mathbf{Y} of Eq. (10) for the complex sound graphed in Fig. 1 with transformation matrix \mathbf{W} from the previous figure. See also Fig. 3. Frequencies are indicated by musical notes on the horizontal axis. It is clear that the calculation has picked up the 2nd harmonic 12 bins (an octave) above the fundamental and the 3rd harmonic 7 bins (a musical fifth) above that for each of these independent components.

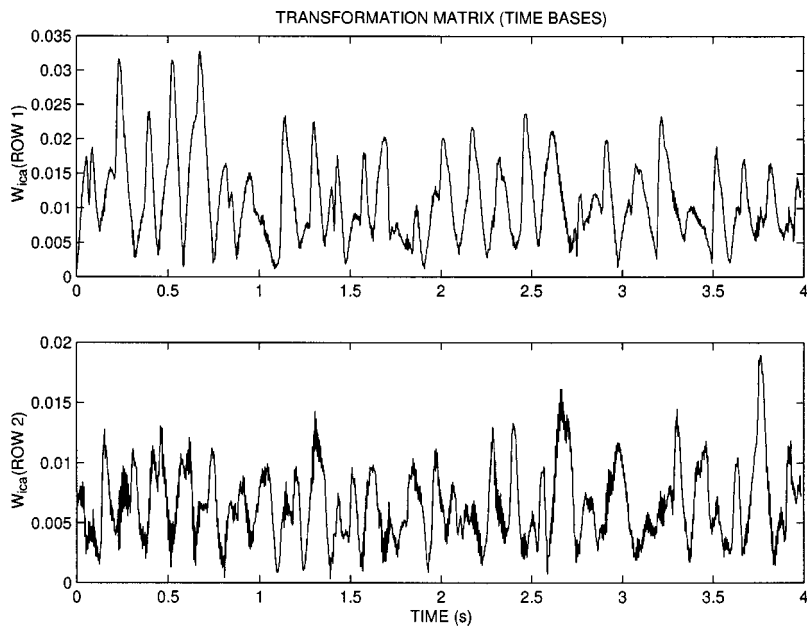


FIG. 8. ICA transformation matrix—the two most important rows from the matrix \mathbf{W} of Eq. (10) for the constant-Q transform of the computer-driven Yamaha disklavier.

5 Hz and found that over the range 250–1000 Hz, fusion occurs at a difference frequency of roughly 30 Hz. Note that the modulation rate corresponds to a note rate of 10 Hz. The terminology “note rate” is used to avoid confusion with frequency of trill pairs. They conclude that a whole-tone trill (12% frequency difference) will be heard as alternating between two notes for frequencies over 400 Hz and as a warble below 125 Hz. The region between these frequencies is ambiguous and depends on the perception of the individual subject. See Table I for a comparison of these to other background studies.

Performance studies are more directly related to our results. Palmer (1996) found that the number of trills in an ornament depends on the tempo, which implies that the trill rate changes less than might otherwise be expected. Note rates varied from 11 Hz (measurement over 11 trill pairs) in

a slow passage to 13.4 Hz (measurement over 9 trill pairs) in a fast passage.

Moore (1992) states that piano trills require one of the fastest alternating movements of which the hand is capable. He finds the upper limit to be about 12–14 notes/s.

In earlier work, Moore (1988) studied trills performed on a cello. He concluded that the limit on the trill seems to be derived from both the performer and the instrument. He gives no quantitative data, but his graphical data indicate a note rate of approximately 12 Hz.

III. SOUND DATABASE

The sounds analyzed consisted of two-note trills obtained from three sources:

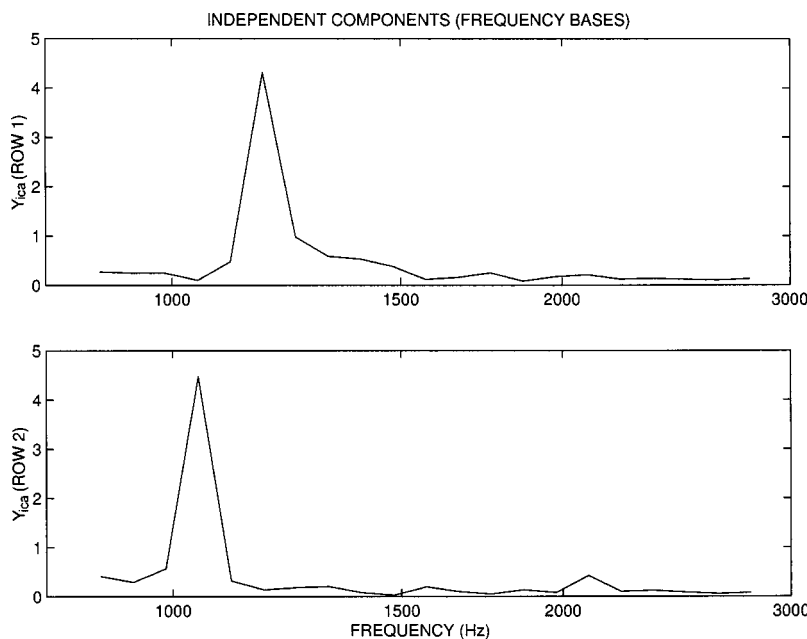


FIG. 9. Independent components—the two most important rows of the matrix \mathbf{Y} of Eq. (10) for the computer-driven Yamaha disklavier with transformation matrix \mathbf{W} from the previous figure.

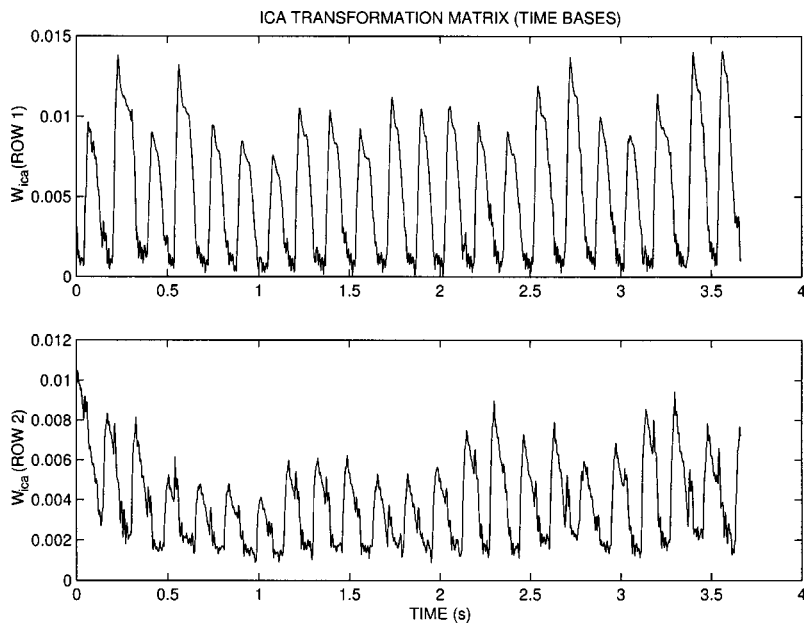


FIG. 10. ICA transformation matrix—the two most important rows from the matrix \mathbf{W} of Eq. (10) for the constant-Q transform of a recorded performance by Charles Fisk. This is an example characterized as “fast with control” by the pianist.

- (1) recordings of a Yamaha Disklavier piano programmed using Miller Puckette’s “pd” program (Puckette, 1996) to drive the piano,
- (2) recordings of pianist Charles Fisk of Wellesley College playing trills on a Steinway S, and
- (3) excerpts from compact disks of performances by Ashkenazy, Horowitz, Goode, Wilde, and Pollini on piano, and Peter-Lukas Graf on the flute.

IV. CALCULATIONS AND RESULTS

Principal component analysis calculations were carried out using Matlab with the function *eig* for diagonalization of a matrix. See Appendix A for details. In our independent component analysis calculations (Appendix B), we used the algorithm Jade¹ and assumed that two notes were present by specifying two independent components in the calculation. If we assume fewer ICs than there are notes actually present,

the independent components will consist of mixtures of the notes. If we assume more ICs than notes actually present, the notes will be evenly distributed across components.

A. Synthetic signal

Using known input as a first example, we compare the results using principal component analysis with those of independent component analysis for the computer-generated signal described in Figs. 1 and 2. Figures 4 and 5 show the quantities \mathbf{U}^T , the transformation matrix, and \mathbf{Y}_{pca} , the principal components, calculated from Eq. (6) and keeping the two most important principal components. The titles of the figures indicate frequency dependence (frequency basis functions) or time dependence (time basis functions).

Looking at the frequency bases of Fig. 5, we find that PCA has picked out the peaks corresponding to the two fundamental frequencies present. These are the dominant fre-

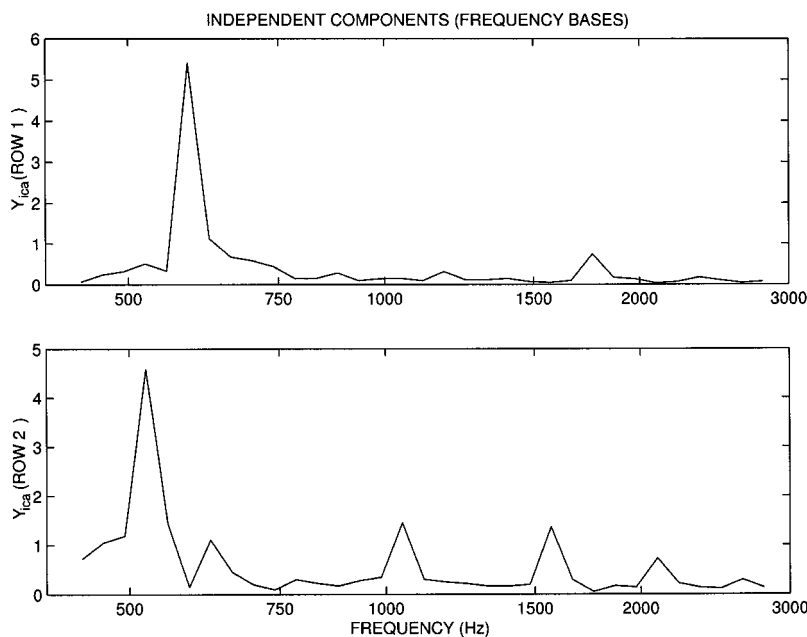


FIG. 11. Independent components—the two most important rows of the matrix \mathbf{Y} of Eq. (10) for the recorded performance of Charles Fisk with transformation matrix \mathbf{W} from the previous figure.

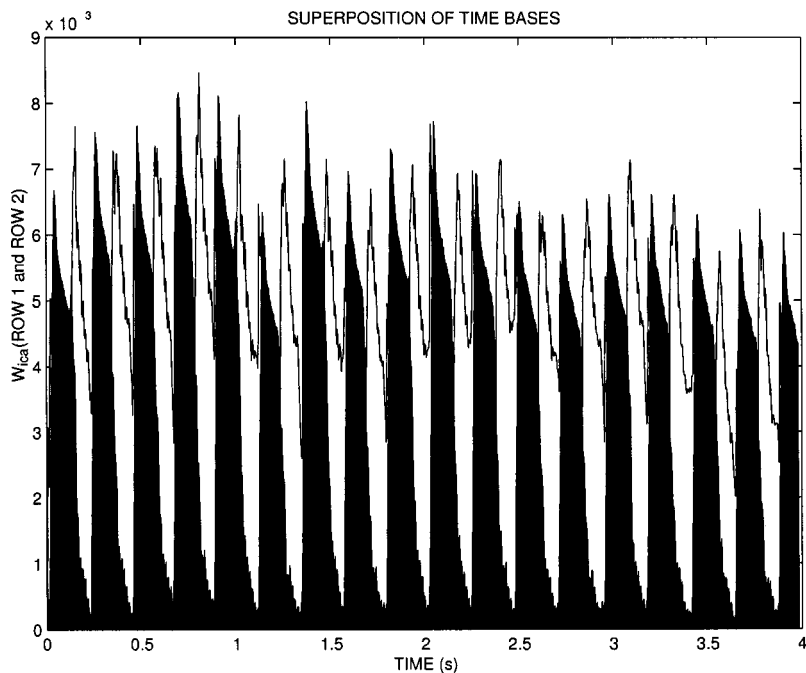


FIG. 12. Superposition of the times bases for one of the “slow” trills recorded by Charles Fisk. This shows clearly the spacing of the notes.

quencies in these data. But in choosing bases, PCA has chosen linear combinations of these two frequencies corresponding to the sum and difference of the two sources rather than separating them. This is a perfectly valid solution for PCA since these are orthogonal bases and are solutions to the eigenvalue equation Eq. (15).

Examining the time bases of Fig. 4 corresponding to these two principal components, they do not contain useful information about the temporal behavior of the two musical notes. The addition and subtraction has effectively removed the possibility of getting times of single note onsets.

Applying the ICA algorithm Jade to the same input (Fig. 2) to obtain W and Y of Eq. (10), the time bases and frequency bases seen in Figs. 6 and 7 are obtained. See also Fig. 3 for the operation applied to the first third of the file. Absolu-

te values were plotted in these and other ICA results. The low-frequency sawtooth modulation of Fig. 6 is an excellent representation of the two alternating sounds simulating a trill, and the two independent components of Fig. 7 are a near-perfect extraction of the frequencies present in each of two complex sounds which were mixed. ICA has thus performed an excellent separation and yielded the two sources which are present while discarding redundant information.

B. Computer-driven piano

To test this method on real sounds, a Yamaha Disklavier piano was driven by computer at a number of different rates with whole-tone trills beginning on the notes C_5 or C_6 . Recordings were made with a Sony TCD-D8 DAT recorder and

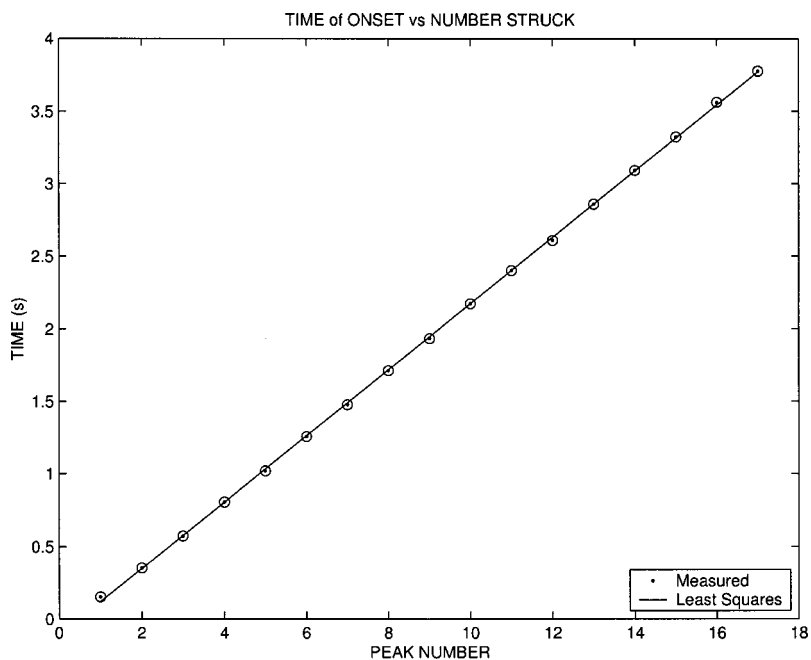


FIG. 13. Onset time against peak number for the peaks of one note of the previous figure compared to a least squares linear fit showing accuracy of note striking.

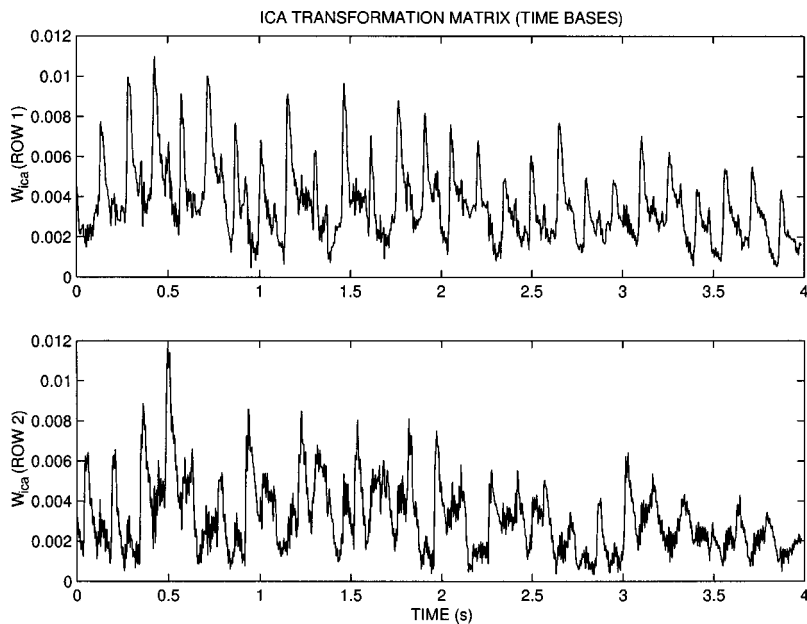


FIG. 14. Transformation matrix—the two most important rows from the matrix W of Eq. (10) for the constant-Q transform of the Pollini performance of Beethoven's Piano Sonata No. 32, Op. 111. This is included as an example of a performance analyzed from CD.

analyzed using the ICA algorithm Jade described previously. The example shown in Fig. 8 has a note rate of 13.5 Hz and is the maximum rate at which this piano could be driven without dropping notes; even so this example is not perfect for the time bases as it is a little beyond the region of reliable operation of the piano. The frequency bases of Fig. 9 are clearly separated, again demonstrating that ICA is able to pull out the relevant information while dropping redundant data.

C. Recordings of live performance

As an example of a live performance, Charles Fiske, a professional pianist and member of the performing faculty of the Wellesley College Music Department, generously agreed to do some trills for this study. In order to determine how a performer views trill rates, he was given the instructions to perform the trills slowly, fast with control, and very fast.

These rates varied from 8.6 for slow to 12.1 notes/s for fast with control (Table I.) ICA results for the time bases and frequency bases are given in Figs. 10 and 11 for one of the “fast with control” examples.

Further analysis was carried out on one of the “slow” files and is shown in Figs. 12 and 13. The superposition of the time bases (black for one note, white for the other) is shown in Fig. 12 in order to demonstrate the precision of the alternating onsets. In a more quantitative graph, Fig. 13 shows the onset times for one of the two notes plotted against note number in order to obtain the average time between trill pairs. This is 0.22 s with a standard deviation of 0.01 showing that the trill is very precise.

D. Examples from compact disk

Trills from a number of performances on compact disk were studied since these had not been previously reported. In

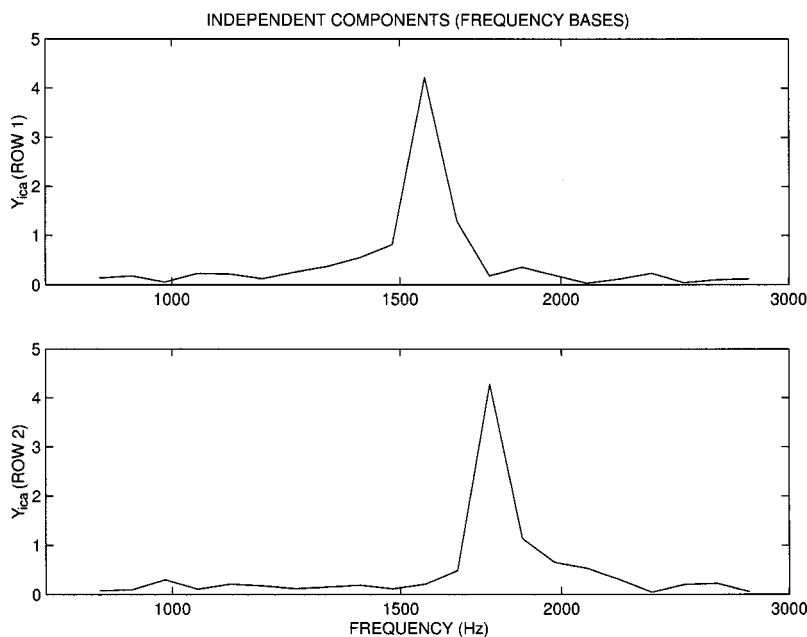


FIG. 15. Independent components—the two most important rows of the matrix Y of Eq. (10) for the Pollini performance from CD with transformation matrix W from the previous figure.

TABLE I. Summary of results on trill rates.

Reference or performer	Notes (or frequency)	Trill rate(note/s)	Comments
Results from Literature			
Shonle/Horan		10, 16	from below and from above
Palmer		13.4, 11	fast passage, slow passage
Moore	D4 E4	13	
		12 to 14	upper limit
Michael Hawley		13	upper limit
Computer-driven piano			
Yamaha 140	C6-D6	14.3	
Yamaha 150	C6-D6	13.5	
Yamaha 170	C6-D6	11.7	
Recording of live performance			
Fisk	C5-D5	12.1	fast with control
Fisk	C5-D5	8.9	slow
Fisk	C6-D6	8.8, 8.6	slow—2 examples
Performances from compact disk			
Pollini		13.5	
Ashkenazy CE 10		16.3	ornament
Ashkenazy MW	A4 B4	11.5	
Goode BA1	D5 E5	13.3	
Goode BA3	F5 G5	15	
Goode BW	G5 A5	13.2	
Horowitz BA1	D5 E5	11.2	
Horowitz BA2	E5f F5	13.8	
Horowitz BA3	F5 G5	13.3	
Horowitz BW	G5 A5	12.6	
Horowitz CE 10	C5 D5	15.7	ornament
Wild CE 10	C5 D5	16	ornament
Flute		12.8	

some cases difficulties in resolving the two notes were encountered due to pedaling, reverberation, or a significant difference in amplitudes of the two notes. Graphs of the transformation matrix and independent components for a particularly good example by Pollini playing Beethoven's Piano Sonata No. 32, Op. 111 are shown in Figs. 14 and 15. This is interesting in that the amplitudes of the two notes are

almost exactly equal (arbitrary units on the vertical axis of Fig. 14), showing great control by the performer.

In order to demonstrate the applicability of this method to instruments other than the piano, our calculation was applied to a flute trill from Mozart's Flute Concerto No. 1. K313. The notes are extremely well resolved as seen in Fig. 16, but the amplitudes are not equal as in the previous ex-

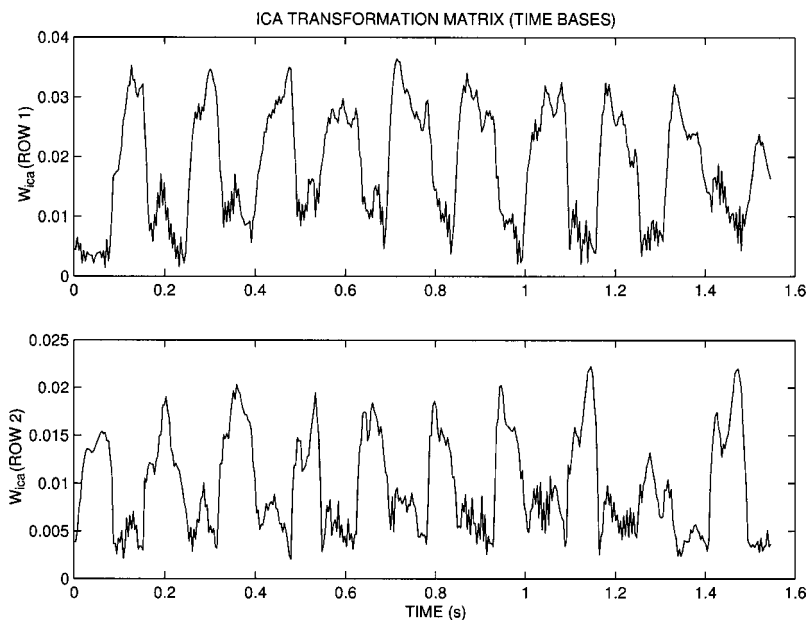


FIG. 16. ICA transformation matrix—the two most important rows of transformation matrix \mathbf{W} of Eq. (10) for the constant-Q transform for the Mozart flute recording.

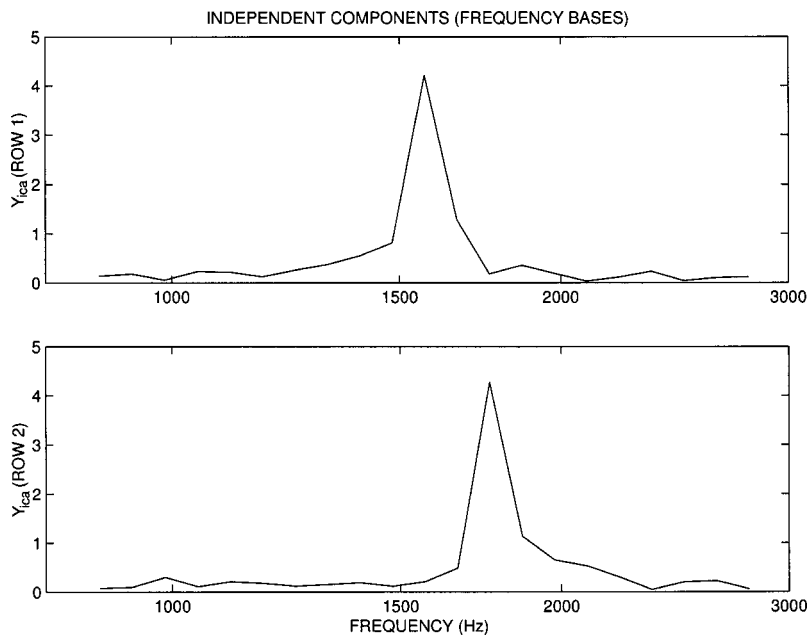


FIG. 17. Independent components—the two most important rows of the matrix \mathbf{Y} of Eq. (10) for the Mozart flute recording with transformation matrix \mathbf{W} from the previous figure.

ample by Pollini. The frequency bases from Fig. 17 show little evidence of higher harmonics indicating that in this frequency range the flute sound is close to a pure tone.

E. Summary of results on trills

Our data on trills is collected in Table I. Most of our results, including the flute trill, are in the range 12–14 notes/s predicted by Moore (1992) and pianist/computer scientist Michael Hawley² in a discussion with one of the authors. Pianist Charles Fisk in the recorded live performance was given the instructions to play slowly, and then fast with control. The “fast with control” example at 12.1 notes/s is consistent with Moore’s and Hawley’s predictions. The ornaments from the Chopin Etude Op. 10, No. 8 played by Ashkenazy, Horowitz, and Wild were all very fast at 16 notes/s, but this was not a sustained trill.

It is interesting to compare performances of the same trill by different performers. The first trill in Beethoven’s Sonata Op. 57 (Appassionata) was played at 13.3 notes/s by Goode compared to 11.2 for Horowitz, which is significantly faster. The third trill in this piece was also significantly faster in the performance by Goode. And finally, a trill from the Beethoven’s Sonata Op. 53 (Waldstein) offers a similar example. Thus there appears to be a consistent difference in the interpretations of these two performers. This opens a fertile area for further research in musical performance.

V. CONCLUSIONS

In this paper we have introduced a new method of musical analysis and applied it to musical trills. Redundancies inherent in the magnitude spectra of trills were identified, and statistical methods were employed to take advantage of this characteristic so as to reveal their basic structure.

The method of independent component analysis can simplify the description of trills as a set of note pairs described by their spectra and corresponding time envelopes. By examination of these pairs we can easily deduce the pitch

and the timing of each note present in the trill. We have also noted how ICA, by employing higher-order statistics and forcing independence, improves the estimate compared to a straightforward application of principal component analysis.

The analysis itself is bootstrapped only to the data presented and devoid of any musical knowledge. In fact, it is a derivative of methods used for auditory scene analysis, which do not assume any previous auditory knowledge.³ This fact allows us to analyze a wide variety of trills and not be constrained or biased by instrument selection, performance, or scale tuning issues. By avoiding the necessity of preprocessing for the extraction of semantically meaningful features, for example pitch or loudness, another advantage is found in a lower burden of computation and complexity.

Finally, we would like to stress the value of redundancy reduction for more complex musical analysis. We have shown how powerful this concept can be for trills; however, it is also applicable to more complex musical segments. In our future work we plan to expand upon this theme and demonstrate how this method can be applied to musical transcription.

APPENDIX A: IMPLEMENTATION OF PRINCIPAL COMPONENT ANALYSIS IN MATLAB

We take as input the matrix X , an N by M matrix in the orientation of Fig. 2. Using the function *eig* to find the unitary transformation U to diagonalize X ,

$$[U, D] = \text{eig}(X * X' / M), \quad (\text{A1})$$

where M is the number of columns of X . The eigenvalue matrix D is ordered by magnitude of its elements from low to high, and U is ordered correspondingly. If two principal components are desired, the last two columns of U will be taken and called the reduced matrix U_r . In matlab notation, $U_r = U(:, N-1:N)$. The transpose of U_r is plotted in the figures, and referred to as row 1 and row 2 in order of importance.

The function Y_{pca} is called the principal components in the figures and defined in Eq. (5) as

$$Y_{pca} = Ur' * X. \quad (A2)$$

APPENDIX B: IMPLEMENTATION OF INDEPENDENT COMPONENT ANALYSIS IN MATLAB

The algorithm *jade* was used for the independent component analyses in the form

$$[A, Y_{ica}] = jade(X, nc), \quad (B1)$$

where X is the matrix defined in Appendix A and nc is the number of components desired. $nc=2$ in the calculations reported.

A is the inverse of the ICA transformation matrix so

$$W = pinv(A) \quad (B2)$$

and

$$Y_{ica} = W * X \quad (B3)$$

were plotted.

¹A number of algorithms for performing independent component analysis are freely available on the internet, such as Jade, Amari, FastICA, and Bell. They can be found using an internet search engine, or more easily from links from Paris Smaragdis' home page. One file was checked with several of these algorithms to ensure that the results were independent of the algorithm used.

²Michael Hawley, personal communication.

³Because of this lack of knowledge many ICA-based algorithms are called "blind." Knowledge accumulated from previous passes is not used and every example is treated as the first and only set of data the algorithm has encountered.

Amari, S., Cichocki, A., and Yang, H. H. (1996). "A New Learning Algorithm for Blind Signal Separation," in *Advances in Neural Information Processing Systems*, edited by D. Touretzky, M. Mozer, and M. Hasselmo (MIT Press, Cambridge, MA).

Bell, A. J., and Sejnowski, T. J. (1995). "An information-maximization approach to blind separation and blind deconvolution," *Neural Comput.* **7**, 1129–1159.

Brown, J. C. (1991). "Calculation of a constant-Q spectral transform" *J. Acoust. Soc. Am.* **89**, 425–434.

Brown, J. C., and Puckette, M. S. (1992). "An efficient algorithm for the calculation of a constant-Q transform," *J. Acoust. Soc. Am.* **92**, 2698–2701.

Brown, J. C., and Smaragdis, P. (2002). "Independent component analysis for onset detection in piano trills," *J. Acoust. Soc. Am.* **111**, 2446.

Cardoso, J. F. (1990). "Eigen-structure of the fourth-order cumulant tensor with application to the blind source separation problem," *Proc. ICASSP*, pp. 2655–2658.

Cardoso, J. F., and Souloumiac, A. (1996). "Jacobi angles for simultaneous diagonalization," *J. Math. Anal. Appl.* **17**(1), 161–164.

Casey, M. A., and Westner, A. (2000). "Separation of Mixed Audio Sources by Independent Subspace Analysis," in *Proceedings of the International Computer Music Conference (ICMC)*, pp. 154–161.

Deco, G., and Obradovic, D. (1996). *An Information-theoretic Approach to Neural Computing* (Springer, New York, 1996).

Hyvarinen, A. (1999). "Survey on Independent Component Analysis," *Neural Comput. Surv.* **2**, 94–128.

Kramer, H. P., and Mathews, M. V. (1956). "A Linear Coding for Transmitting a Set of Correlated Signals," *IRE Trans. Inf. Theory* **IT-2/3**, 41–46.

Moore, G. P. (1992). "Piano trills," *Music Percept.* **9**(3), 351–359.

Moore, G. P., Hary, D., and Naill, R. (1988). "Trills: Some initial observations," *Psychomusicology* **7**, 153–162.

Palmer, C. (1996). "Anatomy of a performance: Sources of musical expression," *Music Percept.* **13**, 433–453.

Puckette, M. S. (1996). "Pure Data," in *Proceedings of the International Computer Music Conference*, San Francisco, International Computer Music Association, pp. 269–272.

Sandell, G. J., and Martens, W. L. (1995). "Perceptual Evaluation of Principal Component-Based Synthesis of Musical Timbres," *J. Audio Eng. Soc.* **43**, 1013–1028.

Shonle, J. I., and Horan, K. E. (1976). "Trill threshold revisited," *J. Acoust. Soc. Am.* **59**, 469–471.

Smaragdis, P. (2001). "Redundancy Reduction for Computational Audition, a Unifying Approach," Ph.D. thesis, Massachusetts Institute of Technology, Media Laboratory, Cambridge, MA.

Stapleton, J. C., and Bass, S. C. (1988). "Synthesis of musical notes based on the Karhunen-Loeve transform," *IEEE Trans. Acoust., Speech, Signal Process.* **ASSP-36**, 305–319.

Therrien, C. W. (1989). *Decision Estimation and Classification* (Wiley, New York).

Torkkola, K. (1999). "Blind separation for audio signals—are we there yet?" *Proc. 1st Int. Workshop Indep. Compon. Anal. Signal Sep.*, Aussois, France, pp. 239–244.