Automated Classification of Piano–Guitar Notes

Dimitrios Fragoulis, *Member, IEEE*, Constantin Papaodysseus, Mihalis Exarhos, George Roussopoulos, Thanasis Panagopoulos, and Dimitrios Kamarotos

***Abstract*—In this paper, a new decisively important factor in both the perceptual and the automated piano–guitar identification process is introduced. This factor is determined by the nontonal spectral content of a note, while it is, in practice, totally independent of the note spectrum tonal part. This conclusion and all related results are based on a number of extended acoustical experiments, performed over the full pitch range of each instrument. The notes have been recorded from six different performers each of whom played a different instrument. Next, a number of powerful criteria for the classification between guitar and piano is proposed. Using these criteria, automated classification between 754 piano and guitar test notes has been achieved with a 100% success rate.**

***Index Terms*—Musical instrument classification, noontonal spectrum, timbre identification, timbre recognition.**

I. Introduction

THE term timbre refers to the attributes of sound that play an important role in the perception of speech and music. According to the American National Standards Institute, timbre is defined as "that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar" [1].

In many of the early studies on instrument timbre, perceptual experiments have been performed to relate acoustic perception with several spectral and temporal characteristics of acoustic signals. Clark *et al.* [2] found that timbre is associated with the attack transient, using modulation during the steady state of a note. Strong and Clark [3], [4] interchanged spectral and temporal envelopes of sounds produced by wind instruments and found that the results are instrument dependent.

In addition, a considerable amount of research has been done in order to find the perceptual dimensions of musical instrument timbre. Grey [5], Grey and Gordon [6], Grey and Moorer [7], used Multidimensional Scaling to put in evidence the main perceptual dimensions of timbre. Similar results were presented by Krumhansl [8] and McAdams [9]. Although many studies have been presented emphasizing the importance of the note onset for instrument identification, Kendall [10] demonstrated that in musical phrases, properties of the steady state are at least as important as transient properties.

The results of the aforementioned studies have been applied to the development of musical instrument identification systems. First attempts on computer identification included a very limited number of instruments and note ranges. Kaminsky and Materka [11] used features derived from the note's energy envelope and a neural network classifier to discriminate guitar, piano, marimba and accordion tones over an one-octave band. De Poli and Tonella [12] used a Self Organizing Map to classify sounds with a procedure similar to Grey's. Cosi *et al.* [13] used features based on an auditory model followed by a neural network to classify instruments. Most of the recent musical instrument identification systems have already shown a respectable level of performance. However, they haven't demonstrated the ability of generalization i.e., the ability of the system to perform successful timbre identification among instrument recordings different from those used during the training procedure. Martin [14], [15] presented a system that operates on single isolated notes played over the full pitch ranges of 15 orchestral instruments and uses a hierarchical classification framework. Recognition was performed using temporal features calculated from the outputs of a log-lag correlogram. Brown [16] used cepstral coefficients calculated from oboe and saxophone samples and managed to develop a hierarchical classifier. Herrera *et al.* [17] have presented a very informative review on the techniques that have been so far proposed for automatic classification of musical instruments.

In this paper we have tackled the problem of piano and guitar timbre determination and classification. We have chosen to deal with these two instruments since their discrimination presents serious difficulties for the following reasons.

- The timbre of those two instruments is quite similar. In many instances even an experienced auditor cannot decide whether the note he is listening to comes from a piano or a guitar.
- Piano and guitar have overlapping frequency ranges and their sound production mechanism is based on string vibration. Moreover, in many instances, both instruments' notes show similar patterns of time decay.
- A fully successful automated classification of piano and guitar timbre has not, so far, been achieved, especially when numerous notes coming from various instruments are considered. Systems presented so far have not demonstrated the ability of generalization. One main reason for this may be the fact that these systems are based on several acoustical characteristics, which are not associated with the kernel of the timbre.
- The question where exactly the timbre lies for both piano and guitar has not been answered so far.

In fact, a series of original experiments is presented that allow for the discrimination of piano and guitar timbre. Based on these experiments, a very important factor in the timbre identification

Manuscript received May 20, 2003; revised December 21, 2004. The Associate Editor coordinating the review of this manuscript and approving it for publication was Dr. Michael Davies.

The authors are with the School of Electrical and Computer Engineering, Division of Communications, Electronics and Information Systems, National Technical University of Athens, GR-15773, Athens, Greece (e-mail: dfrag@mail.ntua.gr).

Digital Object Identifier 10.1109/TSA.2005.857571

1558-7916/$20.00 © 2006 IEEE

process, both perceptual and automated, is introduced, namely the nontonal spectral content of a note. The term "nontonal" part of the spectrum is used to describe the frequency components located between the main peaks.

As it will be shown hereafter, this dimension can provide a number of powerful criteria for the classification of musical instruments. The efficiency of these criteria is demonstrated in the case of piano–guitar discrimination, and a system for the automated classification between piano and guitar is presented that exhibits a success rate of 100%, independently of the choice of the training/test set.

II. Performed Experiments to Study Piano and Guitar Timbre

A. An Experimental Material and Evaluation Group

All related experiments have been performed on 1538 isolated notes sampled at 44 100 Hz over the full pitch range of each instrument. We remind that the fundamental frequencies of piano notes range from 27 Hz to 4160 Hz, while those of guitar notes range from 90 Hz to 990 Hz. From the gathered note samples 612 were isolated guitar notes, while 926 were isolated piano ones. The average note duration is around 1.8 sec. All of them have been recorded from six different performers playing a different instrument each, i.e., six different pianos and six different guitars. About 30% of the guitar notes were generated by plectrum while the rest by a human finger. Recorded notes of half of the performers and half of the instruments have been used as a training set, while those of the other performers playing the remaining instruments as the test set. In this way, a training set has been obtained consisting of 483 piano and 301 guitar sample notes, as well as a test set consisting of 443 piano and 311 guitar sample notes. Notice that a number of piano notes have been produced by the performer so as to bear timbre characteristics, which resemble the guitar ones, and vice-versa. Finally, the acoustic experiments' evaluators were five persons, two professors of musicology and musicians as well, one professional musician and two amateur music lovers.

Recordings were made both in studio and in an ordinary environment, for example a room, a laboratory, a large lecture hall, etc., using digital media. No processing other than the editing of the useless material before and after the sound objects was performed.

B. Timbre Investigation and Determination Experiments

The timbre of an instrument depends on characteristics that appear throughout the whole frequency and time domain. In order to verify this statement, we have performed a number of experiments on the available note samples, with the following results.

1) We have removed the first milliseconds of all recorded piano and guitar notes up to 200 ms and we have observed that the sense of timbre is slightly but not essentially reduced, confirming the results of other researchers [18]–[20].
2) We have created harmonic series corresponding to various fundamental frequencies and we have imposed to them typical guitar and piano envelopes, in both time and frequency domain. The resulted signals acoustically approximated naturally produced notes by guitar and piano, respectively, but not always at a satisfactory level and sometimes poorly. Similar results are obtained if one places the harmonic series' frequencies not at integral multiples of the pitch frequency, but on the partials' exact frequency positions of the naturally produced note, whose envelope is used. We have employed this alternative method especially in the case of piano notes, since it is well known that their partials are not exactly harmonic. This experiment does not aim at synthesizing guitar or piano notes. On the contrary, it demonstrates that the time envelope of a piano or guitar note, alone, does not allow for a perfect recognition of the identity of the instrument.
3) We have proceeded a bit further than experiment No2 by performing the following: Firstly, we have reduced all notes to a common length, say 65 536 samples. Consider, next, a tonal peak of the DFT magnitude $\Phi(\omega)$ of an arbitrary note, located say at frequency ω_0. Moreover consider all spectral information located around ω_0 in the interval $[\omega_o - d_1,\ \omega_o + d_2]$, where $(\omega_o - d_1)$ is the greatest integer smaller than ω_0 such that $\Phi(\omega_0 - d_1) \leq a \cdot \Phi(\omega_0)$, and $(\omega_o + d_2)$ is the smallest integer greater than ω_0 such that $\Phi(\omega_0 + d_2) \leq \mathrm{a} \cdot \Phi(\omega_0)$. We call all this spectral information "the ω_0 lobe". The constant a corresponds to a suitably chosen small value, say $\mathrm{a} = 0.1$.

Subsequently, we have taken numerous pairs of notes of the same pitch, one of a guitar and the other of a piano. For each such pair, we spotted the main peaks of the DFT magnitude $\Phi(\omega)$ and the corresponding lobes. Next, we have created artificial signals by means of the following procedure.

For each peak of the two notes at the same or nearby frequency, we have exchanged the corresponding lobes, namely in the guitar note DFT, we have replaced the guitar lobes with the piano lobes of the same or nearby frequency, and vice versa. In other words, as long as the n-th partial exists both in the guitar and piano notes in hand, we exchange the two n-th lobes' magnitude and phase, independently of their exact peak's position. To accomplish the exchange procedure we use the following rules of thumb.

- Wherever the inserted lobe is of greater width than the removed one, the incomer's values prevail.
- Wherever the inserted lobe is of smaller width than the removed one, then magnitude and phase interpolation is performed to fill the gap.
- In both cases, to avoid abrupt transitions, smoothing is applied in magnitude and phase in a few samples around the points of change.

Next, we have performed the inverse DFT in the obtained spectra. The acoustical impression of the obtained real signal indicated that in some cases there was a slight shift of the "sound color" of the one instrument to the other, while in most cases there was no essential change in the timbre sensation. Therefore, the shape, the relative width and the amplitude of the lobes, by themselves do not allow for a perfect recognition of the piano and guitar identity.

4) The most characteristic experiment was associated with the division of the spectral pattern into a tonal and a nontonal part. This division may be achieved by the method proposed and described below. However this division can be performed by the Spectral Synthesis Method (SMS), too [21], [22]. SMS is an efficient method, but we have preferred the subsequent one since, in some instances, it offers a more accurate representation of the nontonal part. Thus, we have removed all spectral information outside the tonal peaks' lobes of many guitar and piano notes, we have performed inverse DFT on the remaining information and, as a result we noticed that an essential part of the corresponding timbre sensation was lost. In other words, the experiment demonstrates that if one removes the nontonal spectral information, then the color of the note drastically changes and the sensation of the identity of the instrument that has generated it, is lost.

The aforementioned experiments indicate that the associated factors, by themselves do not allow for a perfect recognition of the identity of the two instruments. However, experiment no. 4, together with no. 5, to be presented below, indicate the existence of a new factor very important to piano–guitar timbre perception and automated classification. We will demonstrate that this factor forms the basis of an automated timbre discrimination procedure with 100% success rate, independently of the pitch, the performer and the quality of sound registration, provided that this quality lies above a certain acoustical threshold. In fact, we have performed the following experiment, as well:

5) We have removed the tonal part of the DFT, namely all spectral information lying on the note tonal series peak lobes. In this way, only the nontonal spectral part of the note, say $\mathrm{NT}(\omega)$ remains. The $\mathrm{NT}(\omega)$ energy is a very small fraction of the overall note energy. Next we have taken the inverse DFT of $\mathrm{NT}(\omega)$, which we call *INT*. Subsequently, by acoustically testing the *INT* of all guitar and piano notes we have observed that, in the case of piano, it includes a knock or hammer sound, while in the case of guitar a strumming sound. Notice that *INT* contains an essential part of the attack but it incorporates in addition a serious amount of other instrument related information, as well, including the way the excitation fades out with time. The timbre of these sounds seems to characterize each instrument. Therefore the nontonal spectral part of the notes could be used for the effective discrimination between piano and guitar.

C. Experiments and Characteristics of the Non-Tonal Spectral Part

A typical sound signal of a piano or a guitar note is characterized by a number of partials produced by the harmonic vibration of the instrument strings. However, there are additional nonharmonic modes of vibration that may affect the sound of the instrument. All performed experiments show that the spectral location of the nontonal components is characteristic of piano and guitar instruments. In other words, the performed experiments demonstrate that if one removes the nontonal spectral information, then the color of the note drastically changes and the sensation of the instrument that has generated it, is lost.

More specifically, we have observed the following:

1) In the guitar case, the magnitude of the nontonal spectrum manifests few, usually one, dominant peaks located below 600 Hz.
2) In the piano case on the contrary, the magnitude of the nontonal spectrum is more equally distributed, exhibiting a wider shape, usually with considerably more dominant peaks.

For both piano and guitar, there are several quantitative features of the nontonal spectral magnitude that retain their value in certain intervals characteristic of the instrument, independently of the note pitch. From the acoustical point of view there is a clear discrimination between the two classes of the nontonal content, one corresponding to the guitar and the other to the piano.

III. Nontonal Spectrum Extraction and Class Representatives

The extraction of the nontonal content of a note is performed in the frequency domain, as described below. The DFT is applied to the note signal $x(t)$ and the magnitude of the transformed signal $X(\omega)$ is calculated. Subsequently, all partials are spotted and the amplitudes of the corresponding peaks are estimated. Around each of these peaks, a region is defined so as to include the whole spectral lobe (see Fig. 1). The boundaries of the region correspond to the frequencies for which the spectral magnitude becomes equal to 2% of the peak amplitude of the lobe. By zeroing all these spectral regions that include the tonal spectral lobes of the signal, elimination of the lobes is achieved. In this way a curve is obtained that we call "nontonal spectral pattern" of the note (see Fig. 2). Alternatively, one can obtain the nontonal part of a note zeroing the DFT value of all samples in the interval $[\omega_o - 0.15 \cdot \text{ Pitch } \omega_o + 0.15 \cdot \text{Pitch}]$, where ω_o is the frequency of each DFT peak. Next, in order to reduce the effect of the zero-magnitude DFT points, both linear and by spline interpolation have been used in the nontonal pattern of the note in hand, with similar results. Finally, the interpolated unknown note nontonal pattern $\mathrm{NT}(\omega)$ is smoothed by sliding a moving average L sample window throughout its length with sample step 1, to obtain the smoothed pattern $\mathrm{NTS}(\omega)$. Namely, we have applied throughout $\mathrm{NT}(\omega)$ the operator T_{MA}, obtaining

$$\mathrm{NTS}(\omega) = \mathrm{T_{MA}}(\mathrm{NT}(\omega)) = \frac{1}{L} \sum_{k=-(L-1)/2}^{(L-1)/2} \mathrm{NT}(\omega - k), \quad k \in \mathbf{Z}, \quad L \text{ odd}. \tag{3.1}$$

A good choice of L, although not unique, seems to be $L = 121$, which is independent of the pitch of the note. In this way, the note's "final nontonal pattern" $\mathrm{NTS}(\omega)$ is obtained (see Fig. 3).

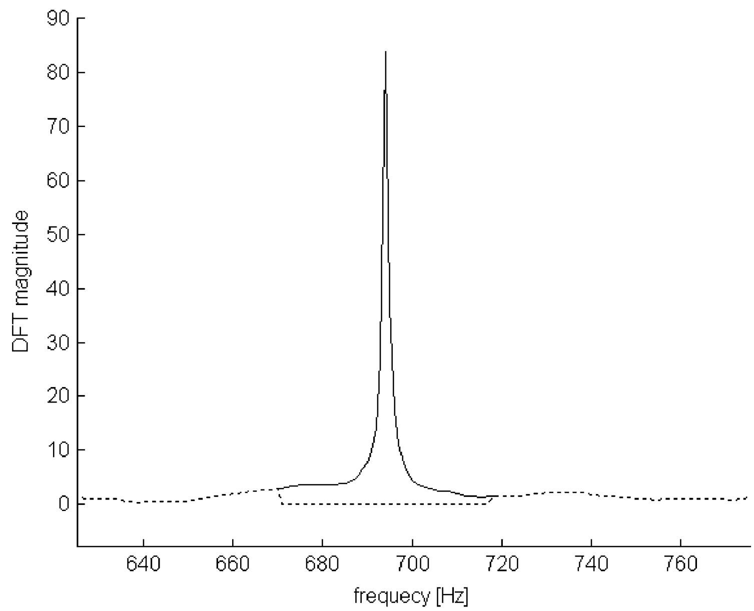


Fig. 1. Plot of an isolated DFT lobe magnitude. *Dotted line*: nontonal DFT part. *Solid line*: DFT lobe.

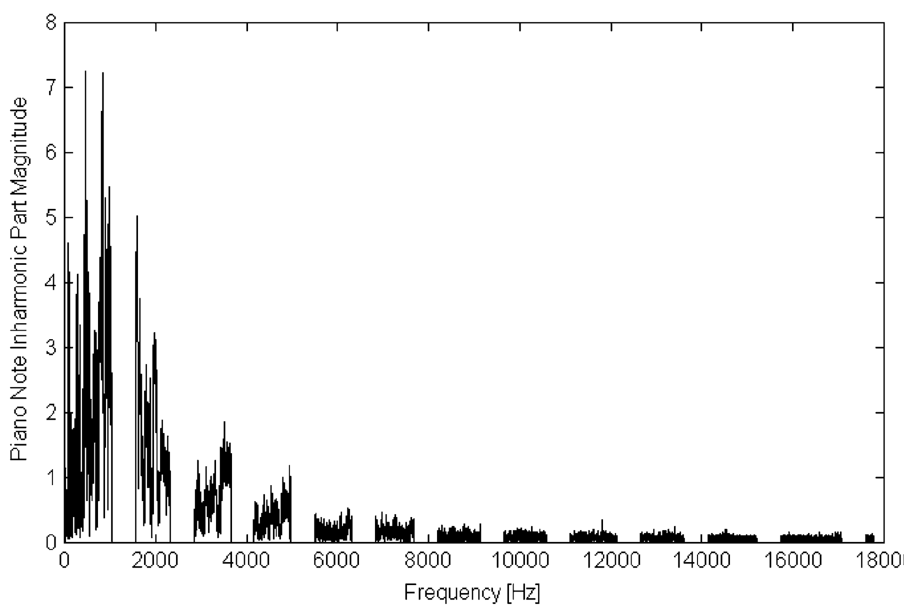


Fig. 2. Typical nontonal spectral pattern of a piano note.

In order to obtain a pattern representing all the nontonal spectral patterns of an instrument, which we call "pattern class representative" or, to put it simply, "class pattern" $\mathrm{CP}(\omega)$, we have applied the following technique.

First, we ensure that all training set notes' DFT's are reduced to the same window length $\mathrm{WL} = 65\,536$ samples. The corresponding pattern $\mathrm{NTS}^i(\omega)$ of all N_T notes of the training set, where $i = 1, 2, \ldots, N_T$, is extracted and normalized to unit energy. For each point of the standard DFT half window of $\mathrm{WL}/2$ length, one computes the mean value of all $\mathrm{NTS}^i(\omega), i = 1, 2, \ldots, N_T$, via

$$\mathrm{CP}(\omega) = \frac{1}{N_T} \sum_{i=1}^{N_T} \mathrm{NTS}^i(\omega). \tag{3.2}$$

In this way, one class pattern Γ_g for guitar and another other class pattern Γ_p for piano are obtained, which are depicted in Fig. 4. This figure shows that these patterns are essentially different.

At this point, in order to test the sensitivity (dependence) of the class pattern in connection with the note pitch, we found the pitch of all available notes of the training set and grouped these notes according to their pitch value. Subsequently, we considered various divisions of the available pitch range into frequency bands and we formed the pattern class representative for each such band, for both instruments. A corresponding example is shown in Fig. 5 and 6, where the whole pitch range is divided into the following three bands.

1) For the guitar: Band 1 [90–166) Hz, Band 2: [167–660) Hz, Band 3: [661–990] Hz.

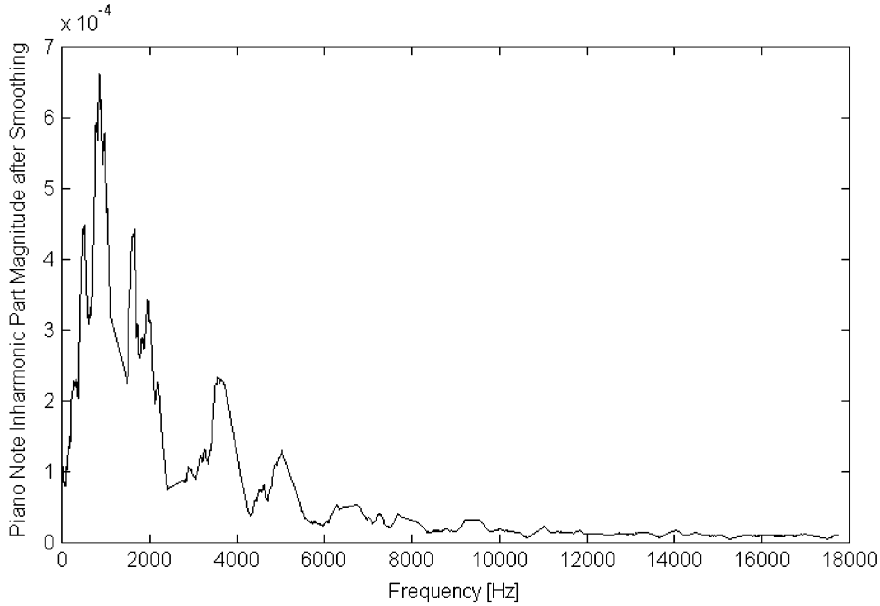


Fig. 3. Final nontonal spectral pattern of Fig. 2 piano note, after interpolation and smoothing.

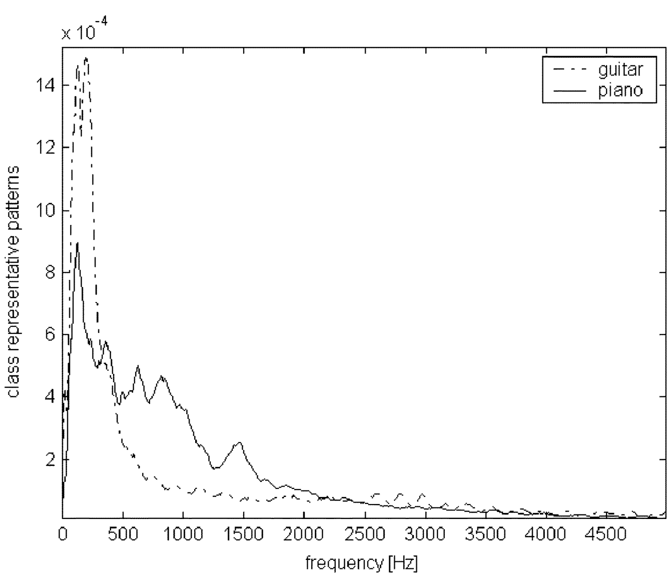


Fig. 4. Estimated class patterns for guitar and piano.

2) For the piano: Band 1: [27–166) Hz, Band 2: [167–660) Hz, Band 3: [661–4160] Hz.

and the pattern class representative has been formed for each such band separately. From these figures, it is inferred that in the guitar case, the representative patterns manifest small variation for all three bands, while in the piano case there is a somewhat greater pattern variation as we move from lower to higher frequencies. However, a careful examination of Fig. 7, where the pattern class representatives for all three frequency bands for both guitar and piano are depicted, shows that there are certain features of these patterns that remain invariant. These features can be used for timbre classification with excellent results, as it will be demonstrated below.

Regarding the sensitivity of these patterns to loudness, we would like to point out that the variation of both Γ_g and Γ_p patterns is too small to prevent a perfect recognition, provided that no distortion (clipping) occurs. Concerning noise, it does not essentially affect the shape of these patterns either, provided it remains below an acoustically acceptable limit. Almost 65% of the available single note samples have been recorded in a natural, nonacoustically isolated environment, e.g., in a room, in a laboratory or in an large lecture hall, where there was a natural background noise.

IV. Extracting Piano–Guitar Discrimination Features and Defining Effective Classification Criteria

Careful observation of the pattern class representatives' shape may suggest where to look for features that can offer decisive classification criteria. First of all, one can divide the whole DFT domain into three regions, according to the pattern class representatives' dissimilarities: In the first region, ranging from 0 to 550 Hz approximately, the guitar representative Γ_g is higher. In the second region, ranging from 550 to 1800 Hz approximately, the piano representative Γ_p is dominant. In the third region, including frequencies greater than 1800 Hz, the guitar representative is slightly higher.

Notice that the guitar class pattern Γ_g demonstrates a high spectral peak and at the same time most of its higher values are located around this peak. On the contrary, the piano class pattern Γ_p values are more uniformly distributed in the first and second region. Based on these observations, the following piano–guitar discrimination features have been extracted.

A. First Classification Feature

If one considers the area enclosed by the final nontonal pattern $\mathrm{NTS}(\omega)$ of a note, which we will name "nontonal" area, a first discrimination feature might be which percentage of this area is found in each of the above regions separately. In fact, for the final nontonal pattern of every recorded note n of the training set, the percentages a_n^1, a_n^2, a_n^3 of the nontonal area corresponding to each of the above-mentioned three regions have been obtained. Subsequently, a simple statistical processing of

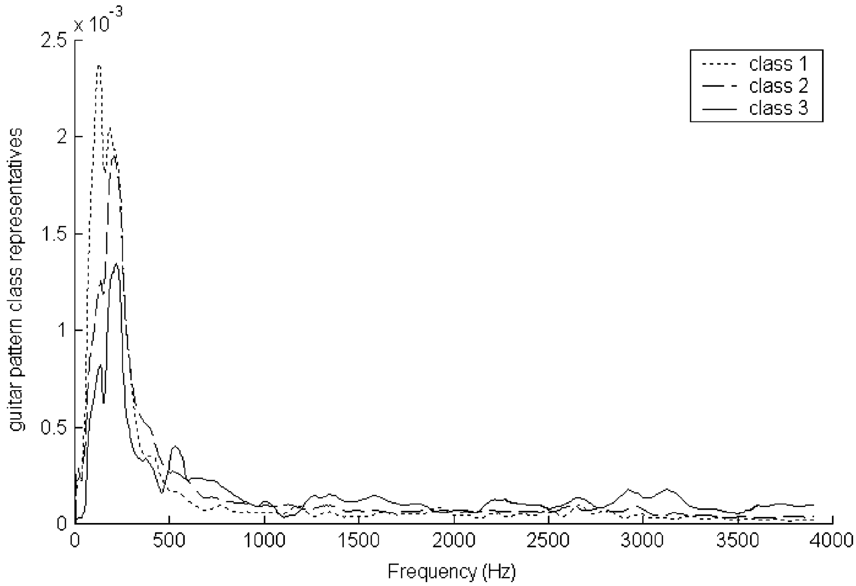


Fig. 5. Guitar class patterns corresponding to the three frequency bands, defined in Section III.

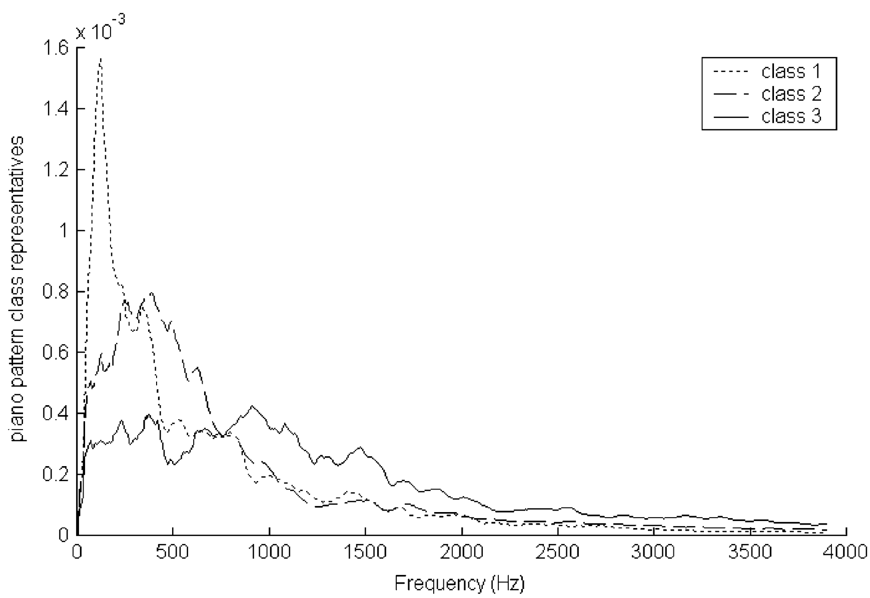


Fig. 6. Piano class patterns corresponding to the three frequency bands, defined in Section III.

these percentages is performed and their histograms are evaluated as follows:

$$a_{\mu}^{i} = \min_{n}\left\{a_n^i\right\}, \quad a_M^i = \max_{n}\left\{a_n^i\right\},$$
$$i = 1, 2, 3, \text{n over all training set sample notes}$$

then one divides all three intervals $[a_{\mu}^{i}, a_M^i]$ into $\mathrm{N} = 25$ equal subintervals and computes all corresponding relative frequencies of appearance a_n^i, thus obtaining the three a_n^i histograms. From this analysis, it follows that the a_n^2 relative frequencies' curves do not overlap, as illustrated in Fig. 8. This result indicates that the percentage of the nontonal area belonging to the second region, namely the one ranging from 550 to 1800 Hz approximately, may constitute a first really good criterion for piano–guitar classification.

We must emphasize that, if the same classification procedure is applied to the magnitude of the whole spectrum instead to its final nontonal pattern only, then the results are dramatically different, in the sense that the relative frequencies' curves (histograms) are highly overlapping, as shown in Fig. 9.

Notice that if one uses the percentage of energy in the second region ρ_2 instead of a_n^2, where

$$\rho_2 = \frac{E_2}{E \cdot \ell_2}$$
$$E_2 = \sum_{\omega \in \text{2nd region}} \text{abs}(\text{NTS}(\omega))$$
$$E = \sum_{\omega \in \text{1st} \cup \text{2nd} \cup \text{3rd region}} \text{abs}(\text{NTS}(\omega))$$

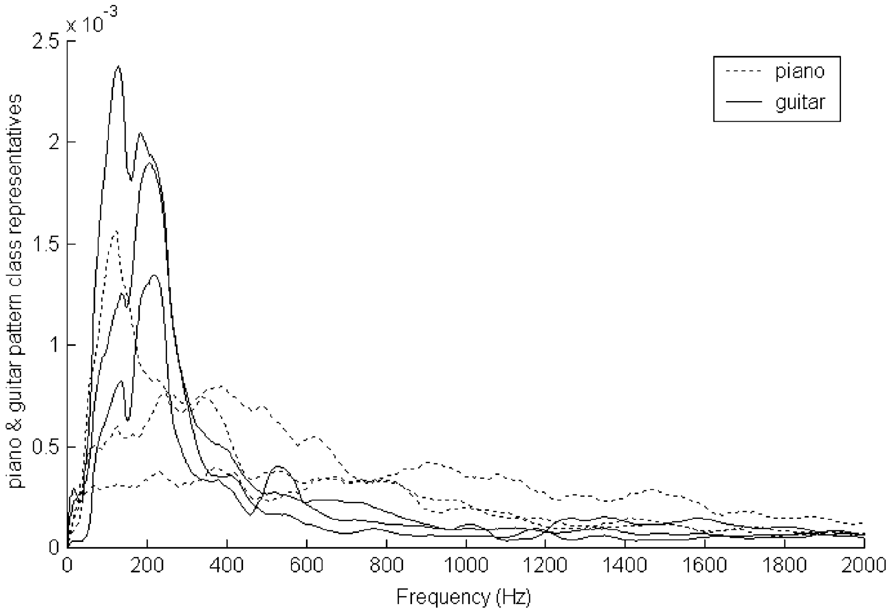


Fig. 7. Comparative demonstration of the guitar and piano class patterns corresponding to the three frequency bands, defined in Section III.

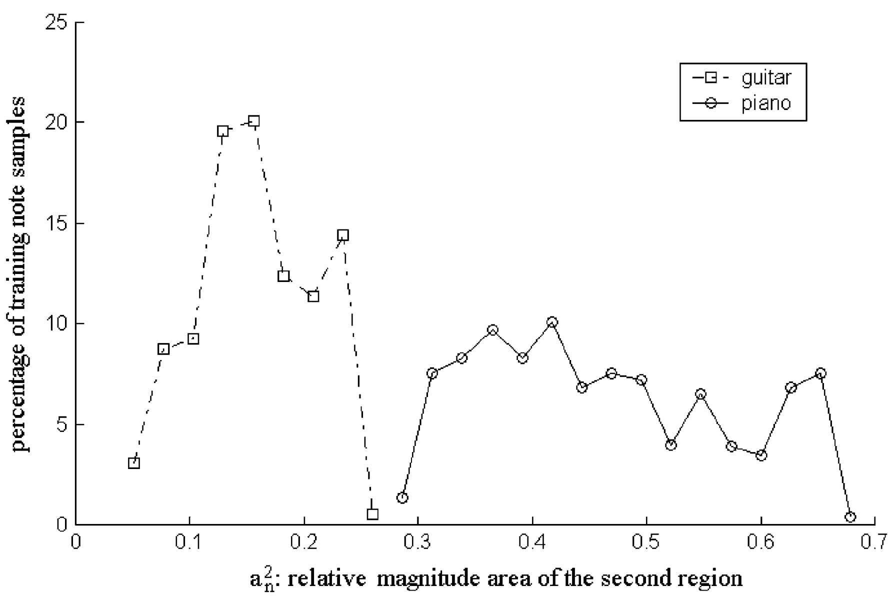


Fig. 8. Nontonal part of the notes: Frequency of appearance (histogram) of the second spectral region relative magnitude a_n^2.

and ℓ_2 the second region length, then the obtained histograms slightly overlap. Thus, the corresponding discrimination feature offers poorer results than the one based on a_n^2.

B. Second Discrimination Feature

Since from Figs. 4 and 7 it is inferred that the class representative patterns of guitar and piano have a quite different magnitude distribution in the first two regions, then the centroid position of the final nontonal pattern may be used as a second discrimination feature.

By obtaining the centroid positions CP^i of the union of the first and second region for all nontonal patterns $\mathrm{NTS}^i(\omega)$, and applying the aforementioned statistical processing to these data, the curves illustrated in Fig. 10 result. The two curves that correspond to piano and guitar, respectively, have a very small overlapping region. Thus, the centroid position of the final nontonal pattern can provide a good criterion for piano–guitar classification.

On the contrary, the centroid position of the whole spectrum generates considerably overlapping frequency curves, as Fig. 11 demonstrates.

C. Third Discrimination Feature

As already mentioned above, the higher values of the guitar class pattern Γ_g are located in a narrow area, while the corresponding piano Γ_p values are more uniformly distributed in the first and second region. In order to quantify this difference in the single notes nontonal spectral patterns of guitar and piano, we have applied the following procedure:

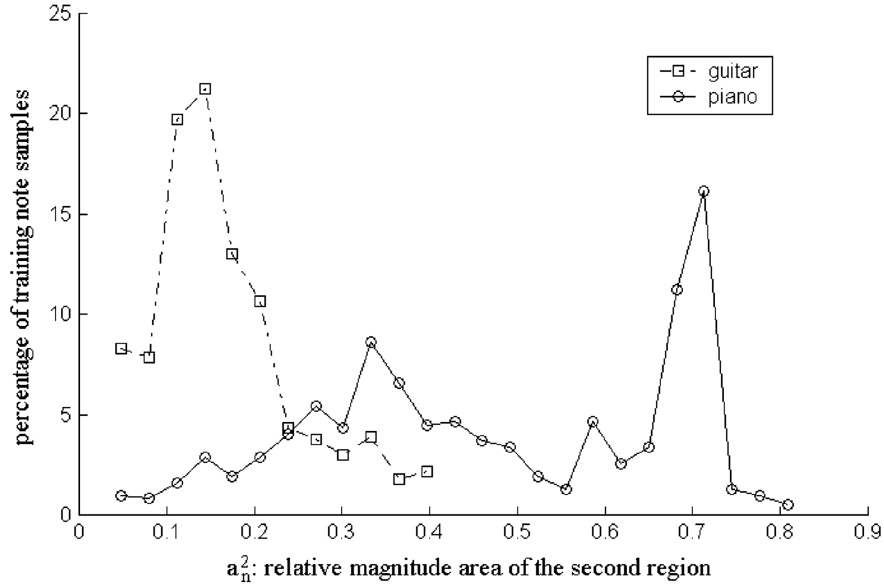


Fig. 9. Total spectrum of the notes: Frequency of appearance (histogram) of the second spectral region relative magnitude a_n^2.

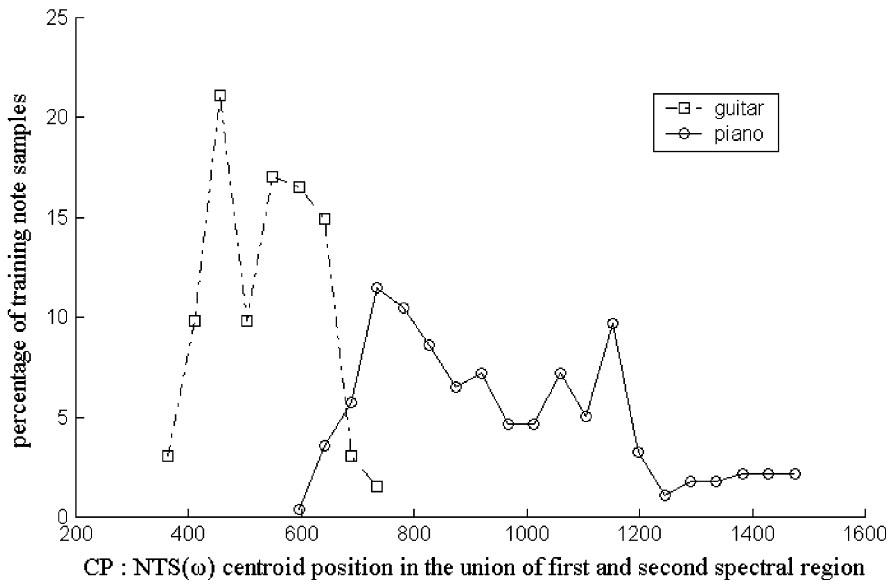


Fig. 10. Nontonal part of the notes: Frequency of appearance (histogram) of the 1st and 2nd spectral region union centroid position CP^i.

We define a suitable threshold Th_3, which is usually chosen to be inversely proportional to the standard DFT window length WL. A good choice seems to be $\mathrm{Th}_3 = (8/\mathrm{WL})$. We then calculate the number of points in the second region where the final nontonal spectral pattern $\mathrm{NTS}^i(\omega)$ is greater than threshold Th_3. The estimated number of points is expressed as a percentage $\mathrm{pnp}_{1,2}$ of the number of points with nonzero nontonal content in the first and second region.

After applying this classification procedure to the training set nontonal spectral patterns, we have obtained a certain distribution of this percentage for the guitar and piano. The corresponding histograms are shown in Fig. 12. From this figure one deduces that this feature may be the basis for a decisive criterion for guitar-piano timbre classification.

It must be stressed once again that, if this classification procedure is applied to the magnitude of the whole spectrum instead to its final nontonal pattern only, then the obtained histograms are highly overlapping as shown in Fig. 13. Therefore, this discrimination procedure, when applied to the whole spectrum, does not offer a decisive guitar-piano classification criterion.

D. Decisive Criteria for Piano–Guitar Classification

Based on the above discrimination features, a number of very efficient piano–guitar classification criteria has been developed. In fact, consider an arbitrary single note generated by either a guitar or a piano. Then, in order to classify it, one first extracts the nontonal pattern $\mathrm{NTS}(\omega)$ as described above and subsequently computes the quantities

1) a_n^2 describing the second region nontonal area;
2) CP, namely the $\mathrm{NTS}(\omega)$ centroid position in the first and second region;
3) $\mathrm{pnp}_{1,2}$ indicating the number of points ω in the first and second region with $\|\mathrm{NTS}(\omega)\| \geq \mathrm{Th}_3$.

The classification of the unknown note can be performed using the following three criteria.

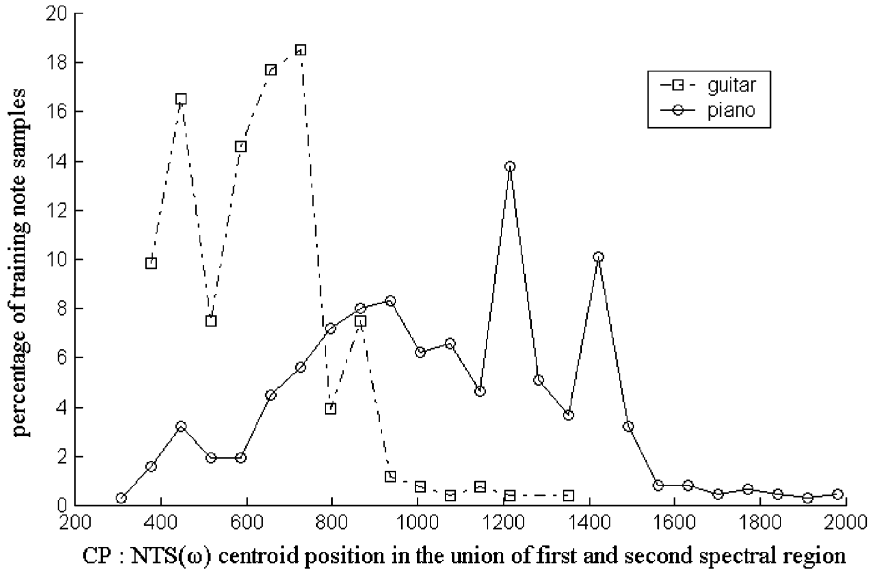


Fig. 11. Total spectrum of the notes: Frequency of appearance (histogram) of the 1st and 2nd spectral region union centroid position CP^i.

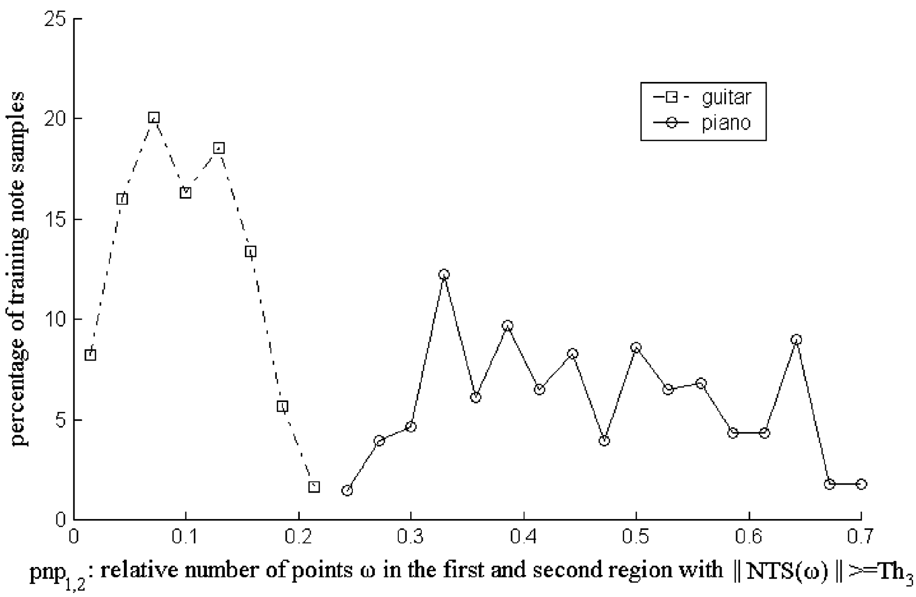


Fig. 12. Nontonal part of the notes: Frequency of appearance (histogram) of the second spectral region relative number of points corresponding to magnitude values over the threshold T_{h3}.

Criterion 1: A tested sample note is classified as a piano note, if $a_n^2 > 0.27$. Otherwise, it is classified as a guitar note.

Criterion 2: A tested sample note is considered to be a guitar note, if the $\mathrm{NTS}(\omega)$ centroid position CP in the first and second region is found below 463 Hz. Otherwise, it is considered to be a piano note.

Criterion 3: A tested sample note is classified as a guitar note, if $\mathrm{pnp}_{1,2} < 0.23$, where $\mathrm{pnp}_{1,2} = \#\{\omega \in Z : 0 \leq \omega \leq 2\pi \cdot 1800 \wedge \|\mathrm{NTS}(\omega \geq \mathrm{Th}_3)\|\}$, (# stands for the cardinal number of the set that follows it). Otherwise, it is classified as a piano note.

We have used the remaining 443 piano and 311 guitar sample notes as a test set and we have applied the aforementioned process on this set, as well as each of the above three criteria separately, for classifying between piano and guitar. The results are presented in Table I. According to the obtained results, each of criteria 1 and 3 is very powerful for classifying between piano and guitar. Criterion 2 is less powerful when compared to 1 and 3. Notice, that criteria 1 and 3 offer the same success rate independently of the choice of the training/test set. However, in multi instrument classification experiments, using the three criteria simultaneously may offer an advantage, since one criterion may fail in the cases where the other two classify the instrument successfully.

V. Conclusion

In this paper, a method for discriminating the piano and guitar timbre is presented. This method employs various techniques for processing the note signals of an instrument, which are followed by corresponding perceptual experiments. In this way, one may alter a number of physical characteristics of a note in the time and/or frequency domain and then one may decide

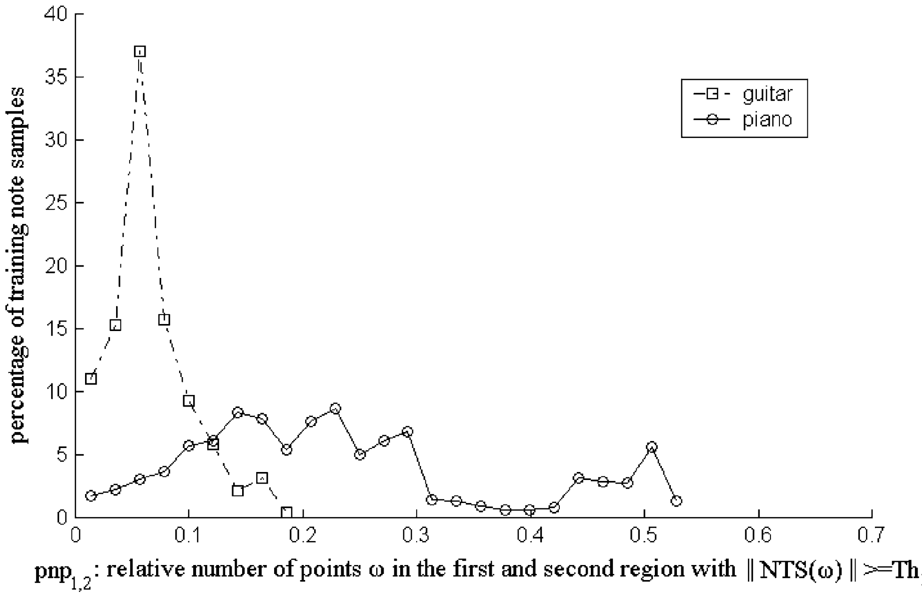


Fig. 13. Total spectrum of the notes: Frequency of appearance (histogram) of the second spectral region relative number of points corresponding to magnitude values over the threshold T_{h3}.

TABLE I
SUCCESS RATE OF PIANO-GUITAR CLASSIFICATION, FOR EACH OF THE THREE SELECTED IDENTIFICATION CRITERIA, SEPARATELY

	Classification success rate among guitar samples	Classification success rate among piano samples	Overall classification success rate
Criterion 1	100%	100%	100%
Criterion 2	95.5%	90.1%	92.6%
Criterion 3	100%	100%	100%

if these characteristics play an essential role to the instrument timbre sensation.

Application of this approach to the piano guitar timbre discrimination led to the original conclusion that a decisively important factor in both the perceptual and the automated instrument identification process is strictly related to the nontonal spectral content of a note. This ascertainment allowed for the development of two powerful criteria for the discrimination between guitar and piano with a 100% success rate independently of the choice of the training/test set. The experiments have been performed on 612 isolated guitar notes and 926 isolated piano notes, over the full pitch range of each instrument.

It seems that the introduced novel methodology can be applied to many other instruments in order to specify the set of those note physical characteristics that bear the decisive part of instrument timbre and for the development of powerful discrimination criteria. In particular, we will investigate the extent to which the information associated to the note nontonal part can be employed to achieve automatic multi-instrument timbre identification. In fact, extensive research is currently carried out toward these directions.

REFERENCES

[1] Psychoacoustical Terminology (1973).
[2] M. Clark, P. Robertson, and D. A. Luce, "A preliminary experiment on the perceptual basis for musical instrument families," *J. Audio Eng. Soc.*, vol. 12, pp. 199–203, 1964.
[3] W. Strong and M. Clark, "Synthesis of wind-instrument tones," *J. Acoust. Soc. Amer.*, vol. 41(A), pp. 39–52, 1967.
[4] ——, "Perturbations of synthetic orchestral wind-instrument tones," *J. Acoust. Soc. Amer.*, vol. 41(B), pp. 277–285, 1967.
[5] J. M. Grey, "Multidimensional perceptual scaling of musical timbres," *J. Acoust. Soc. Amer.*, vol. 61, pp. 1270–1277, 1977.
[6] J. M. Grey and J. W. Gordon, "Perceptual effects of spectral modifications on musical timbres," *J. Acoust. Soc. Amer.*, vol. 63, pp. 1493–1500, 1978.
[7] J. M. Grey and J. A. Moorer, "Perceptual evaluations of synthesized musical instrument tones," *J. Acoust. Soc. Amer.*, vol. 62, pp. 454–462, 1977.
[8] C. L. Krumhansl, "Why is musical timbre so hard to understand?," in *Structure and Perception of Electroacoustic Sound of Music*, S. Nielzen and O. Olsson, Eds. Amsterdam, The Netherlands: Elsevier, 1989, pp. 43–53. Excerpta Medica 846.
[9] S. McAdams, S. Winsberg, S. Donnadieu, G. De Soeteand, and J. Krimphoff, "Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes," *Physiol. Res.*, vol. 58, pp. 177–192, 1995.
[10] R. A. Kendall and E. C. Carterette, "Difference thresholds for timbre related to spectral centroid," *Proc. 1996 4th Int. Conf. Music Perception and Cognition*, pp. 91–95.
[11] I. Kaminsky and A. Materka, "Automatic source identification of monophonic musical instrument sounds," *Proc. 1995 IEEE Int. Conf. Neural Networks*, pp. 189–194.
[12] G. De Poliand and P. Tonella, "Self organizing neural networks and Grey's timbre space," in *Proc. 1993 ICMC*, pp. 441–444.
[13] P. Cosi, G. De Poli, and G. Lauzzana, "Timbre classification by NN and auditory modeling," in *Proc. 1994 Int. Conf. Artificial Neural networks*, pp. 925–928.
[14] K. D. Martin and Y. E. Kim, "Musical instrument identification: A pattern-recognition approach," in *Proc. 136 Meeting Acoustical Society of America*, Oct. 13, 1998.
[15] K. Martin, "Sound—source recognition: A theory and computational model," Ph.D. dissertation, Mass. Inst. Technol., Cambridge, 1999.

[16] J. C. Brown, “Computer identification of musical instruments using pattern recognition with cepstral coefficients as features,” *J. Acoust. Soc. Amer.*, vol. 105, no. 3, pp. 1933–1041, 1999.

[17] P. Herrera-Boyer, X. Amatriain, E. Batlle, and X. Serra, “Toward instrument segmentation for music content description: A critical review of instrument classification techniques,” in *Proc. Int. Symp. Music Information Retrieval, ISMIR 2000*, Plymouth, MA, Oct. 23–25, 2000.

[18] K. W. Berger, “Some factors in the recognition of timbre,” *J. Acoust. Soc. Amer.*, vol. 36, pp. 1888–1891, 1964.

[19] P. Iverson and C. L. Krumhansl, “Measuring the similarity of musical timbre,” *J. Acoust. Soc. Amer.*, vol. 89, p. 1988, 1991.

[20] L. Wedin and G. Goude, “Dimension analysis of the perception of instrumental timbre,” *Scand. J. Psychol.*, vol. 13, pp. 228–240, 1972.

[21] X. Serra and J. O. Smith, “Spectral modeling synthesis: A sound analysis/synthesis system based on deterministic plus stochastic decomposition,” *Comput. Music J.*, vol. 14, no. 4, pp. 12–24, 1990.

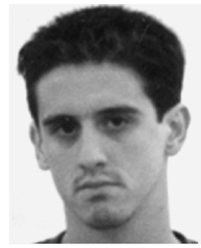
[22] X. Serra, “Musical sound modeling with sinusoids plus noise,” in *Musical Signal Processing*, A. Picciali, C. Roads, and S. Pope, Eds. Amsterdam, The Netherlands: Swets and Zeitlinger, 1997, pp. 5–30.

Dimitrios Fragoulis (M’04) was born in Athens, Greece, in 1973. He received the Diploma and M.Sc. degrees in electrical and computer engineering from the National Technical University of Athens in 1996. He received the Ph.D. degree in computer engineering from the National Technical University of Athens in 2004.

His current research interests concern the application of signal processing techniques to audio signal recognition, the study of the psychological and perceptual aspects of sound, and the application of image processing to archaeology. He has more than ten publications in international journals on these subjects.

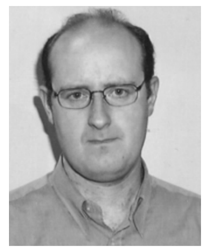
Constantin Papaodysseus was born in Athens, Greece. He received the Diploma degree in electrical and computer engineering from National Technical University of Athens (NTUA) and the M.Sc. degree from Manchester University, Manchester, U.K. He received the Ph.D. degree in computer engineering from NTUA.

From 1996 to 2000, he was an Assistant Professor with the Department of Electrical and Computer Engineering, NTUA. Since 2001, he has been an Associate Professor in the same department of NTUA. His research interests include music and speech processing and automatic recognition, image processing, applied mathematics, algorithm robustness and quantization error analysis, adaptive algorithms, biomedical engineering, etc. He has more than 40 publications in international journals and many publications in international conferences on these subjects.



Mihalis Exarhos was born in Athens, Greece, in 1973. He received the Diploma degree in electrical and computer engineering from Democritus University of Thrace in 1997. He received the Ph.D. degree in computer engineering from NTUA in 2004.

His research interests and recent work are on: image processing, image reconstruction and application of these in archaeology, pattern recognition, sound and music processing, etc. He has five publication in international journals and four publications in international conferences on these subjects.



George Rousopoulos was born in Athens, Greece, in 1971. He received the Diploma degree in computer engineering and informatics from the University of Patras in 1994. He received the Ph.D. degree in computer engineering from the National Technical University of Athens in 2000.

His research interests and recent work are on the following subjects: music and speech processing and automatic recognition, image processing, pattern recognition, algorithm robustness, algorithms for echo cancellation, etc. He has more than ten publications in international journals on these subjects.

Thanasis Panagopoulos was born in Athens, Greece, in 1973. He received the Diploma and M.Sc. degrees in electrical and computer engineering from National Technical University of Athens (NTUA) in 1996. He received the Ph.D. degree in computer engineering from NTUA in 2002.

His research interests and recent work are on the following subjects: music and speech processing and automatic recognition, applications of information theory to archaeology, image processing, pattern recognition, algorithms for echo cancellation, etc. He has more than ten publications in international journals on these subjects.

Dimitrios Kamarotos studied music and computers in Athens, Greece, and continued composition, clarinet and electronic music in Paris, France.

Since 1986, he has been a Research Manager in many projects in collaboration with the National Technical University of Athens, CCMR (Center for Contemporary Music Research). He is combining wind instruments (clarinet, bass-clarinet) and today’s electronic instrumental controllers to enrich human decision during music performance. He has repeatedly collaborated with the Greek National Theater and progressive theater groups. He has particularly worked and published papers on the creation of physically modeled ancient instruments (“Sesclo aulos”). Since 1997, he has been working on automated retrieval of music patterns and timbre recognition.