Determination and Automated Classification of Sax – Flute Timbre

D. FRAGOULIS, M. EXARHOS, C. PAPAODYSSEUS, A. SKEMBRIS,
P. ROUSSOPOULOS, M. PANAGOPOULOS, G. ROUSSOPOULOS
School of Electrical and Computer Engineering
National Technical University of Athens
9 Heroon Polytechneiou, GR-15773, Athens,
GREECE

Abstract: In this paper, a novel approach for saxophone and flute timbre determination and classification is introduced. A set of original experiments including perceptual judgments is presented, that lead to the determination of a minimal ensemble of physical characteristics to which the instrument timbre can be attributed. Using these features, a powerful saxophone - flute timbre discrimination criterion is introduced, offering a 100% success rate between 651 isolated test notes.

Key-Words: musical instrument identification, sax-flute timbre determination, timbre classification

1 Introduction

Given a certain isolated note one can identify the instrument that generated it, by appealing to what is usually called the instrument timbre sensation. Timbre cannot be associated with a single physical characteristic for all instruments and it is distinguished from other attributes such as pitch, intensity and duration. The American National Standards Institute defines timbre 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]. The uncertainty about the notion of timbre is reflected by the huge amount of studies that have tackled this problem.

In many of the early studies on timbre, perceptual experiments have been performed to relate acoustic perception with several spectral and temporal characteristics of acoustic signals [2,3,4,5]. In addition, considerable amount of research has been done in order to find the perceptual dimensions of musical instrument timbre [6,7,8,9]. According to the derived results, timbre is assumed to be a multidimensional quantity, some dimensions of which are related to the spectral and time envelope of the sound.

First attempts in musical instrument recognition operated with a very limited number of instruments and note ranges. Most of the recent musical instrument recognition systems are based on the application of pattern recognition techniques (e.g. statistical methods, neural networks etc.) and have already shown a respectable level of performance [10,11,12,13,14,15,16]. 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 than those used during the training procedure.

In this paper we have tackled the problem of saxophone and flute timbre determination and discrimination. We have focused our study on saxophone and flute, since these two specific instruments seem to produce sounds with a quite similar timbre: in many instances even an experienced auditor cannot decide weather the note he is listening to, comes from a saxophone or a flute. Moreover, a fully successful automated classification of saxophone and flute timbre has been, so far, unsolved and in particular for a large number of note samples. In the following sections, a set of original experiments is introduced, which allow for the determination of saxophone and flute timbre. After spotting the minimal set of physical characteristics that maintains the corresponding instrument timbre, we have defined and applied a powerful saxophone - flute discrimination criterion that exhibits a 100% success rate between 1020 test notes of the two instruments.

2 Performed Experiments -to Spot Sax and Flute Timbre

2.1 Experimental Material and Evaluation Group

All related experiments have been performed on 1181 isolated notes over the full pitch range of each instrument. From the gathered note samples 710 were isolated flute notes, while 471 were isolated

sax ones. All of them have been recorded from four different performers playing a different instrument each. Recorded notes of half of the performers have been used as a training set, while those of the other performers as the test set. In this way, a training set resulted consisting of 212 sax and 318 flute sample notes, as well as a test set consisting of 259 sax and 392 flute sample notes. Finally, the acoustical experiments evaluators were five persons, two professors of musicology and musicians too, one professional musician and two amateur music lovers.

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

2.2 Some necessary definitions

It will be proved necessary for the subsequent analysis to give a number of definitions as follows.

2.2.1 Acceptable Harmonic Peaks Definition

Consider an arbitrary note and its DFT transform performed in a window W of length WL. Due to symmetry, in the following we will use and refer to the half DFT window, say D. Let us consider the sequence of points $P_n \in D$, such that $|F(P_n)| = \max\{|F(i)| : i \in ((n-1/2)\omega_0, (n+1/2)\omega_0)\}$, $n \in N$, where $\omega_0/2\pi$ is the fundamental frequency of the note signal (pitch) and F stands for the discrete Fourier transform of the note signal. We call this sequence of points "the harmonic peak sequence".

As it will become evident from the subsequent analysis, in order that a harmonic peak (partial) contributes to the timbre sensation of a note it should have magnitude greater than a certain threshold T_a. For estimating a good T_a value, we have performed a number of repeated experiments: We have tested a variety of such thresholds and we have accepted the greater one for which no lack of timbre information has been observed. Thus, a respectable value for this threshold is $T_a = 0.2 \cdot E(|F|)$, where $E(|F|)$ stands for the mean value of the DFT magnitude. Hence, any harmonic peak whose amplitude is greater than T_a will be called an acceptable harmonic peak. In the following symbol P_n will stand for the acceptable harmonic peaks only.

2.2.2 'Wide lobe' or simply 'lobe' definition

Let M_n be the mean value of DFT magnitude in the interval $\left(P_n - \omega_0/2, P_n + \omega_0/2\right)$. Next, we define L_n to be the connected subset of Z containing P_n, all frequencies ω_i of which satisfy the inequality $|F(\omega_i)| \geq M_n$, i.e. $L_n = \{\omega_i \in Z : |F(\omega_i)| \geq M_n, P_n \in L_n, L_n \text{ connected}\}$

We will call this subset L_n the wide lobe domain, while $F(L_n)$ is the # n wide lobe or simply # n lobe. It is clear form the above definition that L_n exists wherever acceptable peaks exist. A typical wide lobe of a note signal is illustrated in figure 1.

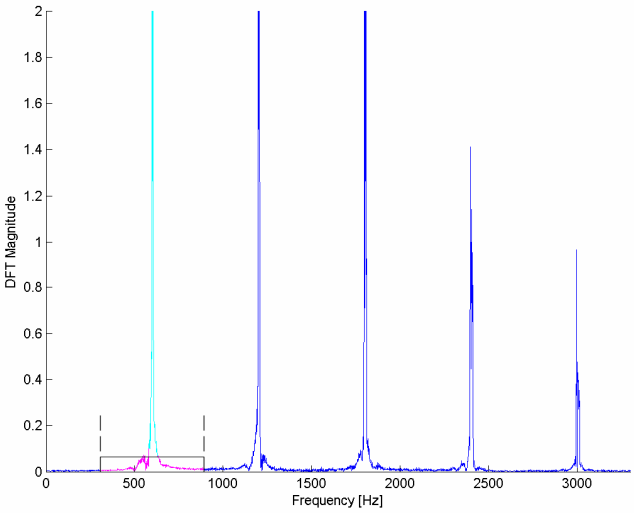

Fig1. A typical wide lobe

2.2.3 Definition of tonal and non-tonal parts

Consider the union of all spotted wide lobes domains, i.e. $H = \bigcup_{all\ n} L_n$; We define $F(H)$ to be the note harmonic or tonal part. We let moreover NH to be the complement of H with respect to the half DFT window D, i.e. $NH = D - H$. $F(NH)$ is defined to be the note non-harmonic or non-tonal part.

The definition of the non harmonic part is essential because the experiments we have performed indicate that the greatest part of piano and guitar timbre lies in NH as it will be described in another paper. However, in the sax and flute timbre case such a statement is not true.

In any case it will be proved necessary for the sax-flute timbre classification and determination to define the notion of the narrow lobe.

2.2.4 'Narrow lobe' definition

Consider once more the sequence of acceptable harmonic peaks P_n and the constant $\gamma = \sqrt[24]{2}$. Now we define the interval $NL_n = [P_n/\gamma, P_n \cdot \gamma]$ and compare it with the wide lobe domain L_n. If $NL_n \subset L_n$ then NL_n is considered to be the narrow lobe domain. Otherwise $NL_n = L_n$ is defined to be so, which means that in this case the meaning of narrow and wide lobe is identical. Extensive experiments have shown that relation $NL_n \subset L_n$ decisively holds. We define the narrow lobe to be $F(NL_n)$. A typical narrow lobe of a note in depicted in figure 2.

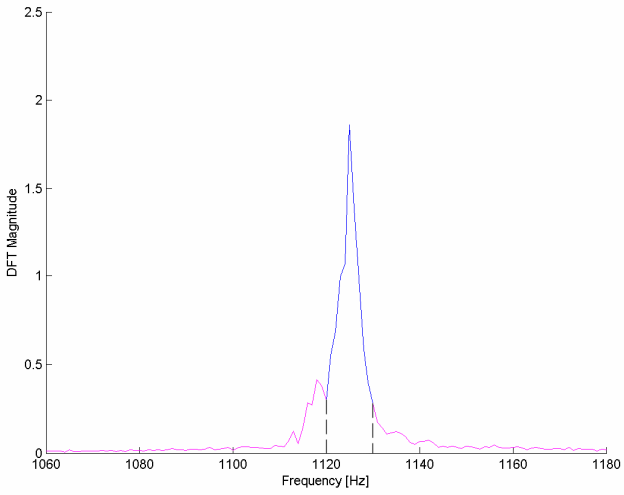

Fig 2. A typical narrow lobe

2.3 Experiments demonstrating where sax-flute timbre can be attributed

In the following a set of experiments will be presented that have been performed on the training set notes. The scope of the experiments is to spot the ensemble of characteristics in the note representation in the time and/or frequency domain, which are decisively responsible for the timbre sensation.

2.3.1 Experiment 1

We consider a sax or flute note as a one dimensional signal, we obtain the envelope of this signal and we try to test if a part of the timbre is seated on this envelope. In order to test this we have proceeded as follows:

We have formed a 'plain' synthetic note $P(t) = \sum_{n=0}^{K} \sin(n\omega_0 t)$, where ω_0 is the fundamental note frequency and K the chosen harmonics number. Subsequently, we have considered an actual note of the test group generated by an instrument O (sax or flute) of the same fundamental frequency ω_0 and we have extracted its time envelope by a successive computation of its maxima and interpolation. In this way a signal $Env(t)$ has been obtained. Next, we have formed the signal $PE(t) = P(t)Env(t)$ for the same time interval. The obtained signal $PE(t)$ has no timbre characteristics at all similar to the ones of the instrument O. In other words, the sax or flute timbre has been completely lost. This experiment indicates that the time envelope of a note seems that does not contain characteristics of the corresponding instrument type.

2.3.2 Experiment 2

We have considered all the wide lobes of each note of the training set and we have multiplied each one of them by a positive real factor, say ran_fact, randomly chosen in the interval [0.5 1.5]. In this way one obtains a perturbed version of the note DFT. On this perturbed version one applies the IDFT and listens to its real part. This process has been repeatedly applied to all training notes lobes for a great variety of ran_fact factors. When we multiplied lobes whose domain Ln was located in an area lower than a certain frequency threshold, no worthwhile timbre modification has taken place. In other words the notes perturbed in this manner in the lower frequency lobes, clearly preserved the timbre of the initial ones. On the contrary, when the higher frequency lobes were perturbed in this way, the timbre sensation was affected and in some cases essentially.

2.3.3 Experiment 3

We have performed the following experiment: For each note separately we have extracted the fundamental frequency ω_0, as well as the first wide lobe domain L_1. Next, we have computed the spectrogram of the note using 64 successive windows of length $N = 4 \cdot 1024$, 75% overlapping, covering the note in the time domain. In each one of these windows we have considered the DFT magnitude maximum in L_1, say $M\Delta_i$ $i = 1, 2, ..., 64$. The graph of $M\Delta_i$ as a function of time index i, essentially displays the evolution of the fundamental partial in time. We have repeated the aforementioned procedure for the second and third harmonic.

Careful examination of the obtained results shows that for a given instrument there is no essential repeatability neither in the rate of change of each one of these harmonics separately nor in the relative onset time. Thus, it seems difficult to associate sax-flute timbre with these characteristics and therefore to employ them in automatic timbre classification.

2.3.4 Experiment 4

Consider an arbitrary pair of notes of the same pitch ω_0, say $N^F_{f_0}, N^S_{f_0}$, where $N^F_{\omega_0}$, has been generated by a flute while $N^S_{\omega_0}$, has been produced by a sax. Now, we have performed the following experiment: If F^F and F^S are the Discrete Fourier Transforms of the two notes $N^F_{f_0}, N^S_{f_0}$ respectively, we have changed these DFTs by altering their L^F_i, L^S_i common number harmonic lobes only via the operation: $L^S_i = \frac{M^F_i}{M^S_i} L^S_i$, $L^F_i = \frac{M^S_i}{M^F_i} L^F_i$. In other words, we have spotted the numbers of the flute and sax wide lobes and we have considered the smaller of these two numbers, say SN. Next we have altered the first SN wide harmonic lobes magnitudes, in both sax and flute so that the magnitude of the saxophone lobes be similar to the magnitude of the flute lobes and vice - versa. In the obtained spectra we have applied the inverse Fourier transform and acoustic experiments on the real part of the signal have been performed. These acoustical experiments indicated that the two instruments timbre had practically remained unaffected.

2.3.5 Experiment 5

Consider the domain L_n of the arbitrary nth lobe consisting of say K points i.e. $L_n = [\lambda_1, \lambda_2, ..., \lambda_K]$. At this point we decrease all lobes by first removing λ_1, then λ_K, then λ_2, next λ_{K-1}, etc. We start decreasing the greater domain amplitude lobe and we continue this process until its length becomes equal to the length of the second greater domain lobe. Subsequently we reduce the previous two lobes domains simultaneously until their length becomes equal to the third greater lobe domain length, and so forth. At each step, the inverse DFT of the obtained reduced tonal spectrum is computed and the resulting timbre sensation is considered. The final conclusion is that:

As far as all reduced lobe domains according to the above procedure remain greater than the corresponding narrow lobe NL_n, then the timbre sensation remains practically unaltered. When some lobe domains L_n are further decreased so as to become subset of NL_n then the timbre sensation is gradually lost.

This series of experiments demonstrates that the narrow lobes sequence essentially contains the necessary information for acoustical saxophone or flute identification.

2.3.6 Experiment 6

From the moment that we have been convinced that the narrow lobes contain the essential flute-sax timbre information we have next proceeded as follows:
We have completely zeroed the lower frequency narrow lobe i.e. $F(NL_1)$. Next we have taken off $F(NL_2)$ i.e. the second lower frequency narrow lobe, etc, and we have acoustically tested the resulting inverse DFTs. These experiments indicated that as far as the removed lobes domains were lower than a certain frequency threshold, then the timbre sensation was not affected. On the contrary, when the lobes removal operation was continued beyond this limit the feeling about the instrument that generated the note was gradually lost.

This series of experiments indicates that the decisive part of the acoustical information leading to saxophone and flute timbre identification could be restricted to the set of high frequency narrow lobes. The exact value of the frequency threshold above which critical timbre information exists, seems to be pitch dependent. A simple empirical rule for determining a sufficient threshold value TC is the following:

One computes the maximum magnitude of the acceptable peaks $MP = \max\{|F(P_n)|\}$ and defines the following quantity $TM = 9 \cdot 10^{-3} \cdot MP$. Subsequently we spot the first narrow lobe, say the μ^{th} one, satisfying $|F(NL_\mu)| \le TM$. The lower point of NL_μ is a respectable representation of the sought threshold TC.

2.4 A very powerful criterion for automated sax- flute discrimination

After the analysis introduced above concerning the training set notes and the related results, we have defined a very powerful sax – flute identification criterion. In fact, in order to determine this criterion

we have first applied the following procedure to all training set notes:

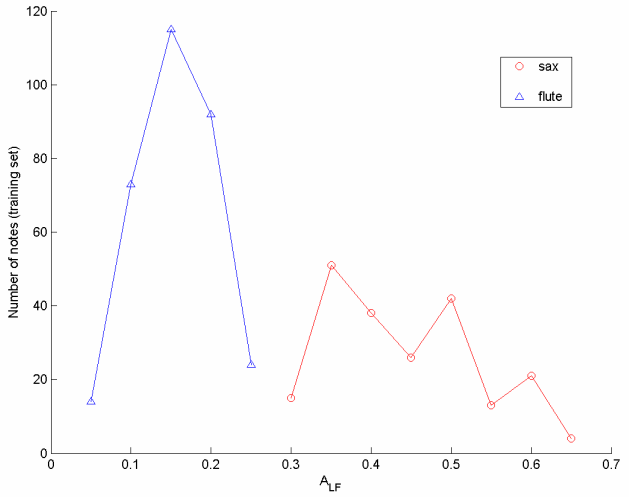

Fig 3. A_{LF} value of the training set notes

Given an arbitrary note we perform the subsequent operations on it:

1. We compute the note DFT, say $F(\omega)$.
2. We spot the note DFTs acceptable peaks.
3. We spot the wide lobes.
4. We divide the note DFT to its harmonic (tonal) and non-harmonic (non-tonal) part, $F(H)$ and $F(NH)$ respectively, by applying the previously introduced method.
5. We keep only the tonal part and reduce the wide lobes domain until it becomes equal to the narrow lobes domain.
6. We dynamically define the frequency threshold TC, as described previously.
7. We keep only the narrow lobes whose domain is entirely greater than TC. We use for this narrow lobes sequence the symbol NLH_i, $i \in N$.
8. We compute the remaining high frequency narrow lobes energy, i.e. $EN(F(NLH_i))$. In addition we compute the energy $EN(F_{TC})$ where $F_{TC} = F(\{\omega : \omega > TC\})$.
9. We calculate the ratio of the high frequency narrow lobes energy $EN(F(NLH_i))$ with the entire high frequency energy $EN(F_{TC})$, namely $A_{LF} = \dfrac{EN(F(NLH_i))}{EN(F_{TC})}$.
10. After applying this procedure to all training set notes we have observed that all flute notes had a A_{LF} value smaller than 0.27 and thus they form a concrete group, while all saxophone notes had a value greater than 0.29 and therefore form another separate group (see figure 3).
11. Next, the following criterion for sax – flute timbre discrimination has been defined and applied to all evaluation (test) set notes:
12. Given an arbitrary evaluation note, we perform to it all aforementioned steps 1-9. If the resulting A_{LF} value is smaller than 0.28 then this note is classified as flute note, otherwise it is classified as a saxophone note.
13. Application of this criterion to all available 351 notes of the test set, offers a completely successful (100%) sax - flute discrimination.

3 Conclusion

In this paper we have introduced a set of original experiments including both acoustical perception and signal processing to determine where saxophone and flute timbre lies. These experiments indicate that the two instruments timbre may be attributed to the higher frequency narrow lobes as they are defined in the paper. Using this minimal ensemble of features pertaining timbre, a very powerful saxophone – flute automated discrimination criterion is introduced. Applying this criterion to 651 isolated test notes, a 100% identification success rate has been achieved.

The set of experiments presented here and in general the whole approach is currently applied to other instruments, too, in order to test if the proposed methodology can effectively treat a multi-instrument identification problem.

References:

[1] *Psychoacoustical terminology*, ANSI Standard S3. 20 – 1973.

[2] M. Clark, P. Robertson and D. A. Luce, A preliminary experiment on the perceptual basis for musical instrument families, *Journal of Audio Engineering Society*, Vol. 12, 1964, pp. 199-203.

[3] J. M. Grey, Multidimensional perceptual scaling of musical timbres, *Journal of the Acoustical Society of America*, Vol. 61, 1967, pp. 1270-1277.

[4] W. Strong and M. Clark, Synthesis of wind-instrument tones, *Journal of the Acoustical Society of America*, Vol. 41(A), 1967, pp. 39-52.

[5] W. Strong and M. Clark, Perturbations of synthetic orchestral wind-instrument tones, *Journal of the Acoustical Society of America*, Vol. 41(B),1967, pp. 277-285.

[6] J. M. Grey and J. W. Gordon, Perceptual effects of spectral modifications on musical timbres, *Journal of the Acoustical Society of America*, Vol. 63, 1978, pp. 1493-1500.

[7] J. M. Grey and J. A. Moorer, Perceptual evaluations of synthesized musical instrument tones, *Journal of the Acoustical Society of America*, Vol. 62, 1977, pp. 454-462.

[8] C. L. Krumhansl, Why is musical timbre so hard to understand? in *Structure and perception of Electroacoustic Sound of Music*, edited by S. Nielzen and O. Olsson, *Excerpta Medica* 846. Amsterdam: Elsevier,1989, pp. 43-53.

[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, *Physiological Research*, Vol. 58, 1995, pp. 177-192.

[10] J.C. Brown, Computer identification of musical instruments using pattern recognition with cepstral coefficients as features, *Journal of the Acoustical Society of America*, vol. 105(3), 1999, pp. 1933-1041.

[11] P. Cosi, G. De Poli and G. Lauzzana, Timbre classification by NN and auditory modeling, *Proceedings of the 1994 International Conference on Artificial neural networks*, pp. 925-928.

[12] G. De Poli and and P. Tonella, Self organizing neural networks and Grey's timbre space, *Proceedings of the 1993 ICMC*, pp. 441-444.

[13] I. Kaminsky and A. Materka, Automatic source identification of monophonic musical instrument sounds, *Proceedings of the 1995 IEEE International Conference of Neural Networks*, pp. 189-194.

[14] R. A. Kendall and E. C. Carterette, Difference thresholds for timbre related to spectral centroid, *Proceedings of the 1996 Fourth International Conference on Music Perception and Cognition*, pp. 91-95.

[15] K. D. Martin and Y. E. Kim, Musical instrument identification: A pattern-recognition approach, *136th meeting of the Acoustical Society of America*, October 13,1998.

[16] P. Herrera-Boyer, X. Amatriain, E. Batlle and X. Serra, Towards Instrument Segmentation for Music Content Description: A Critical Review of Instrument Classification Techniques, *International Symposium on Music Information Retrieval ISMIR 2000*, Plymouth, Massachusetts October 23-25.