# INFORMED SOURCE SEPARATION OF ORCHESTRA AND SOLOIST

**Yushen Han**
School of Informatics and Computing
Indiana University Bloomington
yushan@indiana.edu

**Christopher Raphael**
School of Informatics and Computing
Indiana University Bloomington
craphael@indiana.edu

## ABSTRACT

A novel technique of unmasking to repair the degradation in sources separated by spectrogram masking is proposed. Our approach is based on explicit knowledge of the musical audio at note level from a score-audio alignment, which we termed Informed Source Separation (ISS). Such knowledge allows the spectrogram energy to be decomposed into note-based models. We assume that a spectrogram mask for the solo is obtained and focus on the problem of repairing audio resulting from applying the mask. We evaluate the spectrogram as well as the harmonic structure of the music. We either search for unmasked (orchestra) partials of the orchestra to be transposed onto a masked (solo) region or reshape a solo partial with phase and amplitude imputed from unmasked regions. We describe a Kalman smoothing technique to decouple the phase and amplitude of a musical partial that enables the modification to the spectrogram. Audio examples from a piano concerto are available for evaluation.

## 1. INTRODUCTION

We address the "desoloing" problem, in which we attempt to isolate the accompanying instruments in a monaural recording of music for soloist and orchestral accompaniment. The motivation is to produce the audio of the accompaniment part for concertos in the "classical" domain as well as the karaoke in popular music, whereas the ultimate goal is to have the orchestra adapt timing to the live player, a problem we do not discuss there. Nevertheless, the accompanying audio is needed and we offer solutions through our demixing or isolation of the original sources (instruments).

Most past effort in this "source separation" problem treats Blind Source Separation (BSS) problems and assumes little knowledge of the audio content rather than the independence of the sources [1] or relies on general cues of musical sources rather than the content of the sources [2]. In contrast, we assume explicit knowledge in the form of a score match, which establishes a correspondence between the audio data and a symbolic score representation giving the onset times of all musical events. See Figure 1 for an example. Such correspondence, known as "score following" or "alignment", initially introduced and developed by Vercoe and Dannenberg [12] is the foundation of our approach which we termed *Informed Source Separation* (ISS). Other examples in the category of ISS include Dubnov [6].

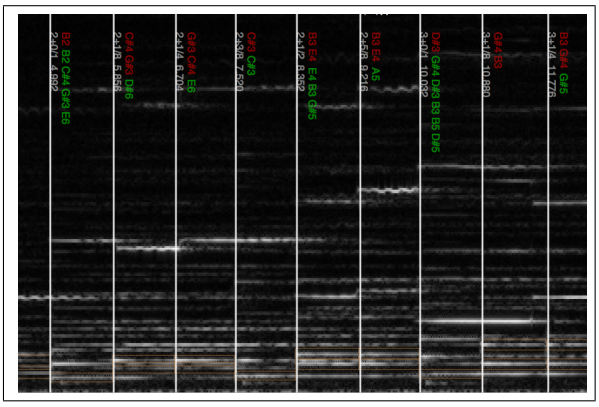

**Figure 1**. Piano note onsets (vertical lines) of an excerpt from 2nd mvmt. of Ravel's piano concerto in G

In our approach, we begin by masking the short time Fourier transform (STFT) in an attempt to "erase" the soloist's contribution. We also based our exploration toward partial-wise phase/amplitude relationship on previous work [7] in which spectrogram magnitude is decomposed into each partial by fitting note-based models. [11] is another effort of spectrogram decomposition in speech. However, our emphasis here is not on estimating the mask or fitting note models, but on employing a novel set of procedures (see sect. 5) that estimates and transforms note partials, in which the damage caused by our masking procedure is repaired.

Here our assumption is that there is information redundancy in terms of phase and amplitude between the "observable" partials (i.e. not significantly overlapped by the solo or an accompaniment instrument of a different family) and damaged partials. Our hope is to "copy and paste" musical partials from the observable area to the damaged area with some necessary transformations that exploit those redundancy to maintain the consistency between the observable and the damaged. These procedures can be automated by analyzing the texture of the music from the score and testing the soundness of remaining partials on the desoloed spectrogram. We call this process *unmasking* in which the masked-out solo regions will be recovered.

The structure of this paper is as follows: we briefly formulate the masking problem in sect. 2, followed by note-based parameterization in sect. 3 and phase estimation in sect. 4. Such estimating enables our repair-by-unmasking technique in sect. 5 which is applied in the context of a piano concerto in sect. 6.

## 2. SPECTROGRAM MASKING OF THE SOLO

Given our original audio signal, $x(s)$, we define the short time Fourier transform (STFT) by

$$X(t,k) = \sum_{n=0}^{N-1} x(tH+n)w(n)e^{-2\pi jkn/N}$$

where $H$ is the hop size, $N$ is the window length and $w$ is the window function. We will define our masking operation in this STFT domain. To do so, we estimate two "complementary" masks, $1_s(t,k)$, and $1_a(t,k)$, taking values in $\{0,1\}$ with $1_s(t,k)+1_a(t,k)=1$. These masks are used to isolate the parts of $X$ we attribute to the soloist and accompaniment through

$$X_s(t,k) = 1_s(t,k)X(t,k) \quad (1)$$

$$X_a(t,k) = 1_a(t,k)X(t,k) \quad (2)$$

In other words we label each time-frequency "cell" $(t,k)$ as either solo or accompaniment. Since our focus here is on the *unmasking* problem, we will bias our labeling of each time-frequency cell toward the solo category, since we want to make sure the original soloist is completely removed. Using our score match, it would be relatively easy to simply draw a rectangle around each solo partial while calling the interior of these rectangles our solo mask. Our approach is somewhat more sophisticated, employing special treatment of the wide spectral dispersion associated with note onsets by Ono et al. [13], as well as careful modeling of the steady state partials. However, we will not discuss this mask estimation problem here.

While $X_a(t,k)$ (and $X_s(t,k)$) is, in general, not the STFT of *any* time signal, applying the inverse STFT operation gives perceptually sufficient results with appropriately defined STFT. In particular, if we use a Hann window with $H = N/4$, one can show that applying the STFT inverse to $X_a$ results in the audio signal whose STFT is closest to $X_a$ in the sense of Euclidean distance [5].

The result of this process elimanates more than the soloist, of course, since the accompanying instruments also contributed to the STFT in the region we have masked out. A possible remedy in sect. 4 is the main focus of our paper.

## 3. NOTE-BASED MUSIC PARAMETERIZATION

In this section we briefly review our parameterization of the music given the score, which is adapted from our technique to decompose the spectrogram magnitude into note models in [7]. This parameterization is also used to facilitate our phase estimation in sect. 4.

From the score, suppose we have a collection of notes $\mathcal{N}$ in the piece of interest, for a note $n \in \mathcal{N}$, we know its instrumentation $i_n \in \mathcal{I}$ where $\mathcal{I}$ is the set of instruments in this piece and can be further partitioned into disjoint subsets $\mathcal{I}_s$ and $\mathcal{I}_a$ for solo and accompaniment instruments separately.

Moreover, we know the time span of note $n$: $T_n = \{t_n^{on}, \ldots, t_n^{off}\}$ from the score following. Also, as the note pitch $p_n$ indicates its set of valid harmonics under a certain Nyquist frequency: $\mathcal{H}_n = \{1, \ldots, H_n\}$, we confine the frequency bin span of each partial $h \in \mathcal{H}_n$ to $K_{n,h} = \{k_{n,h}^{low}, \ldots, k_{n,h}^{high}\}$. $K_{n,h}$ implements a band-pass filter to specify a frequency bin span where the contribution from the partial of interest (very likely to be mixed with other partials of close frequencies) is significant in terms of spectrogram magnitude while the spectral energy outside of $K_{n,h}$ is ignored.

Such 2-dimensional, rectangular time-bin support $B_{n,h} = \{(t,k)|t \in T_n, k \in K_{n,h}\}$ specifies a band-passed filter bank over $T_n$ to extract time domain partial $p_h(s)$ from $X(t,k)$ We denote $B_n = B_{n,1} \cup \ldots \cup B_{n,H_n}$ to be the support for all harmonic components of note $n$.

We then assume a Normal mixture model for the spectrogram magnitude of an orchestra note $n$: each harmonic of the note is one Gaussian component in the mixture with normalized weight $\nu_{n,h}$, coupled frequency bin expectation $\mu_{n,h}(t) = h\mu_{n,1}(t)$, and unknown variance $\sigma_{n,h}^2$. To accommodate the (possibly dramatic) change in amplitude over time of a note, we also introduce a normalized non-negative profile, $\eta_{n,h}(t)$, to outline the frame-wise amplitude of $h$th partial of $n$th note.

Strictly, the centroid of each partial may not be precisely coupled by $\mu_{n,h}(t) = h\mu_{n,1}(t)$. But it is approximately true for all the instruments except for piano in our study. To summarize:

- a weight $\nu_{n,h} > 0$ for $\forall(n,h)$ with $\sum_{h\in\mathcal{H}_n} \nu_{n,h} = 1$
- a time support $T_n = \{t_n^{on}, \ldots, t_n^{off}\}$, which is shared among all partials of note $n$
- an amplitude envelope $\eta_{n,h}(t) > 0$ for $\forall(n,h)$ with $\sum_{h\in\mathcal{H}_n} \eta_{n,h}(t) = 1$
- a frequency bin support $K_{n,h} = \{k_{n,h}^{low}, \ldots, k_{n,h}^{high}\}$
- a frequency bin centroid $\mu_{n,h}(t)$ which reflected the frequency of partial $h$ at $t$. Among different partials, they are coupled by $\mu_{n,1}(t) = \frac{\mu_{n,h}(t)}{h}$
- a frequency bin variance $\sigma_{n,h}$ that describes magnitude distribution of partial $h$ over frequency bins with expectation $\mu_{n,h}(t)$ under Normal assumption.

Finally we can define a"template" function $q_{n,h}(t,k)$

$$= \begin{cases} 0, \ \forall(t,k) : t \notin T_n \ or \ \ k \notin K_{n,h} \\ \nu_{n,h}\eta_{n,h}(t)f(k;\mu_{n,h},\sigma_{n,h}^2); \ \ otherwise \end{cases} \quad (3)$$

where $f(k;\mu_{n,h},\sigma_{n,h}^2)$ is the normal density function. This parameterization is subjected to normalization to ensure $\sum_{h\in\mathcal{H}_n}\sum_{(t,k)\in B_{n,h}} q_{n,h}(t,k) = 1$ for note $n$.

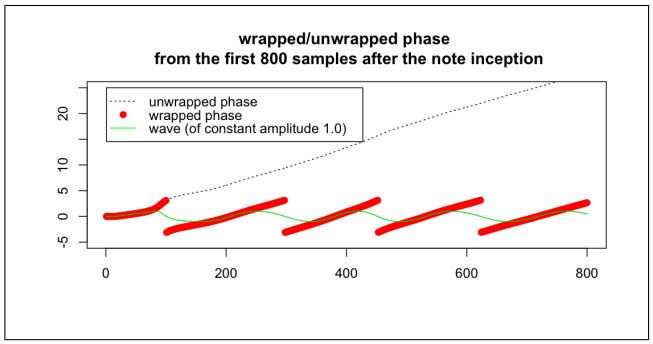


**Figure 2**. Wrapped and Unwrapped Phase

Our assumption is that the magnitude contribution from each note partial indexed by $(n, h)$ to the spectrogram is raised from a collection of independent Poisson random variables $\{Z_n(t,k)\}$ for $(t,k) \in B_n$ [3]. The expectation of $Z_n(t,k)$ is $\delta_n \sum_h q_{n,h}(t,k)$ where $\delta_n$ describes the degree to which $Z_n(t,k)$ contributes to $X(t,k)$. Intuitively, $\delta_n$ is our estimate of the total spectrogram magnitude contribution from note $n$ and can be interpreted as the overall "amplitude" of the note $n$. The estimation of $\delta_n$ is, no doubt, a significant factor of source separation quality and our solution by an EM algorithm is documented in [7]. For the rest of the paper, we assume a somewhat reliable $\delta_n$ is known so we can focus on the unknown phase of each partial of note $n$.

## 4. PARTIAL-WISE PHASE ESTIMATION AND TRANSFORMATIONS

As usually only a subset of partials of a note is damaged by removing the solo partial, we hope to exploited the harmonicity assumption in wind and string instruments supported by Fletcher [9] and Brown [10] to impute the phase of those missing partials in the orchestra. To do so, we first introduce a generic method to decouple the phase and slow-changing amplitude of a band-limited signal in 4.1 which enables our two major tools to "unmask" the damaged spectrogram: harmonic transposition in 4.2 and phase-locked modulation in 4.3.

### 4.1 Phase Estimation by Kalman Smoothing

In this section we represent our note partial, $p_h(s)$, in terms of a time-varying amplitude and phase:

$$p_h(s) \approx \alpha_h(s) \cos(\theta_h(s))$$

where the time-varying amplitude, $\alpha_h(s)$, is non-negative and varies slowly compared with $p_h(s)$, and the "unwrapped" phase (see Figure 2) function, $\theta_h(s)$, is monotonically non-decreasing. A more precise review of the slow-changing $\alpha_h(s)$ in a sinusoidal model is given by Rodet [14].

In order to estimate $\alpha_h(s)$ and $\theta_h(s)$ we follow the model of Taylan Cemgil [8] and view the harmonic, $p_h(s)$, as the ouput of a Kalman filter model [16] [17]. To this end we define a sequence of two-dimensional state vectors $\{x(s) = (x_1(s), x_2(s))^t\}$ where $x_1(0)$ and $x_2(0)$ are independent 0-mean random variables with variance $\gamma^2$, and the remaining variables follow evolution equation $x(s+1) = Ax(s) + w(s)$ where $\{w(s)\}$ is an independent sequence of 0-mean 2-dimensional vectors with independent components of fixed variance (the variance can be tuned empirically). $A$ is the rotation matrix, defined in terms of the expected phase advance per sample, $\rho$, which is directly computable from the nominal frequency of the partial:

$$A = \begin{pmatrix} \cos\rho & \sin\rho \\ -\sin\rho & \cos\rho \end{pmatrix}$$

Thus, $x(s)$ is a sequence of vectors that circle around the origin and an approximately known frequency with variable distance from the origin. We then model our observed partial as $p_h(s) = x_1(s) + v(s)$ where $\{v(s)\}$ is another sequence of independent 0-mean variables with a certain variance (this variance is tuned empirically too).

It is well known that the Kalman filter allows straightforward computation of the conditional distribution, $p(x(s)|\{p_h(s')\})$, and that this distribution is Normal for each value of $s$. Thus we estimate $x(s)$ by $\hat{x}(s) = E(x(s)|\{p_h(s')\})$. The representation of the partial in terms of amplitude and non-decreasing phase follows from the polar coordinate representation of $\hat{x}(s)$:

$$\begin{aligned} \alpha_h(s) &= \sqrt{\hat{x}_1^2(s) + \hat{x}_2^2(s)} \\ \theta_h(s) &= 2\pi k(s) + \tan^{-1}(\frac{\hat{x}_2(s)}{\hat{x}_1(s)}) \end{aligned}$$

where each $k(s)$ is chosen to be the non-negative minimal integer value that ensures that $\theta_h(s)$ is non-decreasing.

Note that for phase sequence $\theta_h(s), s \in \{1, \ldots, S\}$, not only the final phase estimate $\hat{\theta}_h(S)$ but also all previous phase estimates are of interest. To get the "best" phase estimation, we need to update the state estimates backward to incorporate the observation that were not "available" at sample $s$ in the forward pass. This motivates Kalman smoothing (see chapter 5 of [17]) which calculates the smoothed phase estimate $\hat{\theta}_h(s)$ recursively backward from the last sample at $S$.

### 4.2 Harmonic Transposition

With amplitude $\alpha_h(s)$ and phase $\theta_h(s)$ decoupled from $h$th harmonic of a note, we are ready to "project" one harmonic into a different harmonic while maintaining the harmonicity between the source and the destination. Supposing we estimated the unwrapped phase of the $i$th harmonic as $\theta_i(s)$, the "projected" phase sequence at $j$th harmonic is given by $\tilde{\theta}_j(s) = \frac{j\theta_i(s)}{i}$ and the resulting $j$th harmonic by

$$\tilde{p}_j(s) = \tilde{\alpha}_j(s) \cos(\frac{j\theta_i(s)}{i}) \quad (4)$$

where $\tilde{\alpha}_j(s)$ is either known or imputed amplitude at $j$th harmonic. In this work, we usually have an estimate of $\tilde{\alpha}_j(s)$ by scaling $\delta_n$ from sect. 3.

Our harmonic transposition exploit such "harmonicity" between partials, which is a well-studied phenomenon. Early

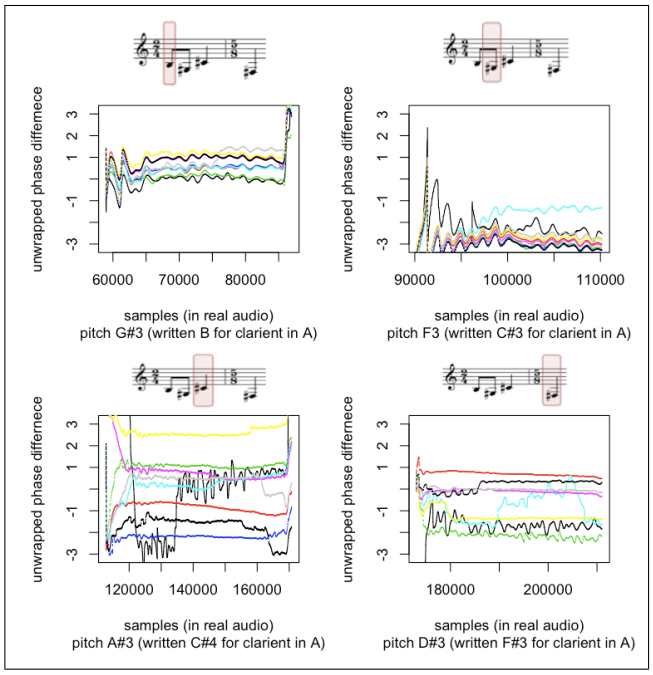


**Figure 3**. Unwrapped Phase Difference

work mainly by Fletcher showed that frequencies of the partials in"the middle portion of the tone" of string instrument are integral multiples of the fundamental frequency by using sonograph and also derived that partials of string and wind instrument are "rigorously locked into harmonic relationship" [9]. By using single frame approximation on a variety of digital samples, Brown concluded that "continuously driven instruments such as the bowed strings, winds, and voice have phase-locked frequency components with frequencies in the ratio of integers to within the currently achievable measurement accuracy of about 0.2%" [10] from experiments with and without vibrato.

To demonstrate such harmonicity in our framework, we focus on the "projection" of the unwrapped phase $\theta_i(s)$ from partial $i$ to partial $j$ by

$$\theta_{i,j}(s) = \frac{j\theta_{h_1}(s)}{i} \tag{5}$$

By "projecting" the phase of different partials to a common harmonic, we can examine such phase relation on a variety of orchestra instruments. We can visualize *pairwise phase difference* $\theta_{i,1}(s) - \theta_{j,1}(s)$ at the fundamental for any $i \neq j$. Figure 3 shows the *pairwise phase difference* for the first 4 notes from a performance of the first movement of Stravinsky's Three Pieces for Clarinet Solo. The salient message from this plot is: the *pairwise phase difference* is in a very small range (mostly $(-\frac{\pi}{2}, \frac{\pi}{2})$) and never drifts away over the entire note; the error (including measurement error and true difference) is not accumulative. This supports our approximation of phase coherence.

Piano and other impulsively driven instruments such as strings played pizzicato are counter-examples whose partials deviate from integer ratios due to the stiffness of the string [10].

### 4.3 Phase-locked Modulation

In addition to the partial-wise relationship, we want to exploit timewise similarity in terms of phase and amplitude within one note.

Suppose we have a partition $T_1 = \{s_1, \ldots, s_k - 1\}$, $T_2 = \{s_k, \ldots, s_2\}$ on the sample indices $T = \{s_1, \ldots, s_2\}$ of the sustaining part of a reasonably long orchestra note, we can only observe the unwrapped phase sequence at $\theta_h(T_1)$ but $\theta_h(T_2)$ is missing. We can impute $\theta_h(T_2)$ sequentially by

$$\theta_h(s_k+n) = \theta_h(s_k+n-1)+\theta_h(s_1+1+n)-\theta_h(s_1+n) \tag{6}$$

for any $0 \leq n \leq s_2 - s_k$. We omit the formula to obtain $\theta_h(T_1)$ if we observe $\theta_h(T_2)$.

This operation preserves the phase advance per sample in $T_1$ and applies such $\Delta\theta_h(T_1)$ cyclically to $T_2$. This is similar to the phase vocoder except for that we are doing it on the sample level rather than frame level. For a long enough time span $T_1$, we are capturing the pattern of frequency fluctuation in $\theta_h(T_1)$. To synthesize a segment of a partial, we also need the amplitude envelope over $T_2$. A simple solution is to reuse the average amplitude $\alpha_h$ over $T_1$ (with some minor modulation) to "sustain" a note through the end of $T_2$. If the orchestra note is holding for quite long, which is common in some orchestration, we are effectively synthesizing the sustaining part of the partial.

## 5. SPECTROGRAM UNMASKING

In an attempt to fix the damage caused by desolo, we examine the spectrogram with a focus on areas where the accompaniment notes (harmonics) are damaged.

In the type of music that we (and many solo musicians) are mainly interested in, for instance, a piano concerto, it is common that a string section may double the solo instrument at the unison, fifth, or octave in either direction. In these cases, masking out the solo part usually results in many damaged partials in the orchestra since consonant intervals mean more partials are likely to share the same frequencies. With this in mind, we use some heuristics to create an algorithm to automatically perform the two partial-wise transformations developed in 4.2 and 4.3. Since the texture of the music can be highly complex, we reconstruct a somewhat "generic" scenario for illustration of this algorithm in Figure 4. The 1-bar score in the figure is a reduction from a piano concerto where the piano part is frequently doubled by the lower string sections.

Supposing we have obtained solo mask $1_s(t,k)$, a damaged region $B^d_{n,h} \subseteq B_{n,h}$, a template $g_{n,h}(t,k)$ and an amplitude estimate $\delta_n$ from section 2 and 3 for a damaged partial $h$ of note $n$, we summarize our *heuristic* algorithm:

First, we need to evaluate the damage. If $\sum_{(t,k)\in B^d_{n,h}} g_{n,h}(t,k) \ll \sum_{(t,k)\in B_{n,h}} g_{n,h}(t,k)$, we leave it as intact; otherwise we need to repair it. Specially, if undamaged part $B_{n,h} \setminus B^d_{n,h}$ is a narrow band-limited "strip" (e.g. a single frequency bin), we need to "expand" the solo mask to remove those initially deemed

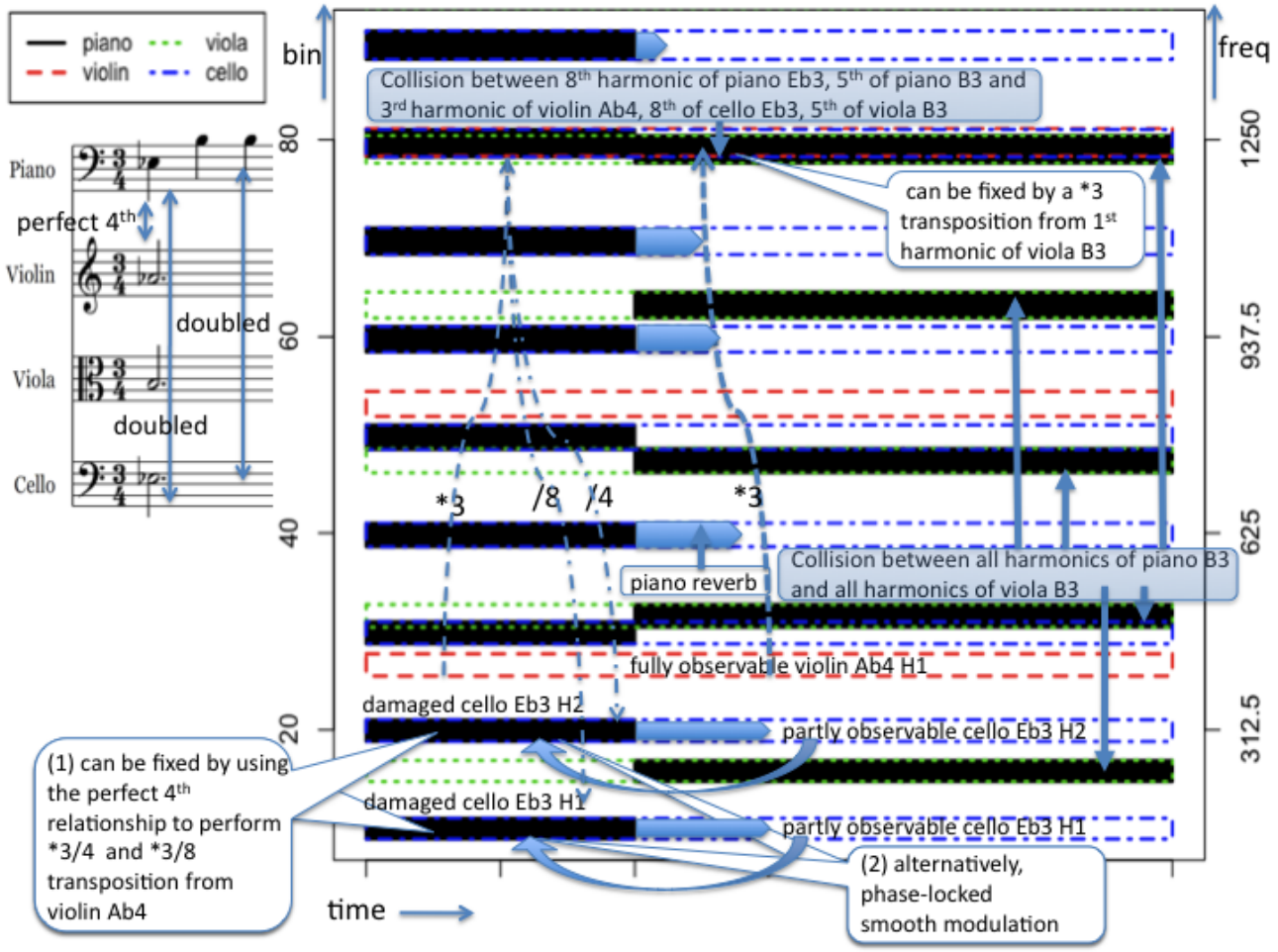


**Figure 4**. Evaluating Desolo Damage and Possible Fix Using Both Score and Spectrogram

"undamaged" f-t cells as well because such residue tends to create artifact "musical noise" whose suppression deserves treatment, mostly from speech enhancement. After such extra "masking", we use $B^u_{n,h} \subseteq B_{n,h}$ to denote the remaining undamaged region.

Second, since $B_{n_1,h_1} \cap B_{n_2,h_2} \neq \emptyset, n_1 \neq n_2$ for possibly many different note partials contributing energy to the same region, we choose one damaged orchestra partial $(n,h)$ to repair: $\underset{(n,h)}{\operatorname{argmax}} \sum_{(t,k)\in B_{n,h}} \delta_n g_{n,h}(t,k)$ assuming Max-Approximation that only one signal dominates in each time-frequency cell [3].

Third, in the score we look for consonant intervals such as octaves, perfect 5th and perfect 4th in the hope to find an observable partial whose frequency is in a relatively simple ratio to the damaged one waiting to be "transposed" to. We call this partial, if exists, a *candidate*. Usually more than one candidate exist. Large modulus value, simple frequency ratio and identical instrumentation are factors that we favor in choosing the best candidate without creating artifacts. Thus, harmonic transposition can be performed vertically on the spectrogram (e.g. from 3rd to 5th harmonic of viola note B3 in Figure 4) if the duration of the candidate partial covers that of the damaged area.

Fourth, when there is no candidate partial for the partial indexed by $(n,h)$, if there exists a partial $(m,i)$ whose time support of its undamaged portion $T^u_{m,i}$ is adjacent to the damaged duration $T^d_{n,h}$ and whose frequency bin support $K_{m,i}$ satisfies $K^d_{n,h} \subseteq K_{m,i}$ we can perform phase-locked modulation with differenced phase sequence estimated from $B^u_{m,i}$ to $B^d_{n,h}$. The 2 cello partials in Figure 4 are repaired this way.

Occasionally, we are unable to perform either transformation and label the damaged partial as such.

## 6. EXPERIMENT RESULTS

We experiment with an excerpt of 45 seconds from the 2nd movement of Ravel's piano concerto in G major.

Table 1 lists a breakdown of the number of partials [1] and the number of harmonic transpositions and phase-locked modulation that our algorithm performed. The last column, "unable to fix" gives the number of occurrences that no undamaged orchestra partial is available to estimate phase from. We relax on that the 4 sections of string instruments can be used to repair each other by harmonic transposition but do not allow any harmonic transposition between two different instruments in the woodwind family. This is be-

[1] the number of partials only include partials that have significant spectral energy and are below Nyquist frequency at SR=8000Hz.

| | note | partial | tran. from | tran. to | modulation | unable to repair |
|---|---|---|---|---|---|---|
| oboe | 20 | 85 | 1 | 1 | 0 | 1 |
| clarinet | 6 | 18 | 3 | 3 | 0 | 0 |
| flute | 6 | 18 | 0 | 0 | 0 | 0 |
| violin1 | 5 | 42 | 14 | 9 | 0 | 0 |
| violin2 | 11 | 107 | 34 | 24 | 24 | 2 |
| viola | 16 | 160 | 33 | 41 | 64 | 5 |
| cello | 12 | 120 | 43 | 50 | 22 | 6 |

**Table 1**. Instrument breakdown of partials being repaired

cause the oboe is sharper than the other two in this excerpt. At the end the most of damaged partials are fixed in some way. We also notice that the woodwinds are less damaged because the notes are very high pitched and too loud to yield to the solo piano at their time-frequency region, while the lower string instruments are frequently damaged.

The original, desoloed-but-unrepaired and repaired audio are available at our demo website `http://xavier.informatics.indiana.edu/~yushan/ISMIR2010` to evaluate the solo mask and improvement from unmasking. Plots in color giving a breakdown of the partials on the spectrogram are also available.

## 7. CONCLUSION, EVALUATION AND FUTURE WORK

Instead of merely extracting one source (instrument) of sound from the mixture, we distinguish our proposed ISS method from other known source separation methods by our explicit *repair* stage that addresses the audio degradation caused by the separation procedure. This stage significantly enhances the perceptual audio quality and boosts performance measurement such as distortion due to interferences proposed by Vincent. That the reconstructed note sounds plausible for some orchestra instruments suggests that the partial-wise phase/amplitude relationship is a potentially fruitful topic to investigate.

At this stage, we admit that the comparison of our method of "unmasking" with other missing data inference techniques such as [15] is not available and hence is our future work. An ideal evaluation of any method of solo/orchestra separation requires a "ground truth" of the two sources recorded separately and an artificial mix of the two. However, such "ground truth" is almost away absent in the real case and the evaluation is mainly subjective. Our exploration begins with a music sample library to artificially construct ground truth according to the score while maintaining the texture of the music of interests.

## 8. REFERENCES


[1] Bell, A. J., and Sejnowski, T. J.: "An Information-Maximization Approach to Blind Separation and Blind Deconvolutionm *Neural Computation*, vol. 7, no. 6, pp. 1129 - 1159, 1995.

[2] E. Vincent: "Musical Source Separation Using Time-Frequency Source Priors," *IEEE Trans. on Speech and Audio Processing* , Vol. 14, Iss. 1, Jan. 2006 pp. 91 - 98.

[3] D. Ellis: Chap. 4 of *Computational Auditory Scene Analysis: Principles, Algorithms and Applications*, Wiley/IEEE Press, pp.115-146, 2006.

[4] A. S. Bregman: *Auditory scene analysis*. MIT Press: Cambridge, MA, 1990.

[5] Francis R. Bach and Michael I. Jordan: "Blind one-microphone speech separation: A spectral learning approach.", *NIPS*, pages 6572, 2005.

[6] S. Dubnov: "Optimal filtering of an instrument sound in a mixed recording using harmonic model and score alignment," *ICMC* 2004, Miami.

[7] Y. Han, C. Raphael: "Desoloing Monaural Audio Using Mixture Models," *ISMIR*, Vienna, 2007

[8] A. T. Cemgil, S. J. Godsill: "Probabilistic Phase Vocoder and its application to Interpolation of Missing Values in Audio Signals." Antalya/Turkey, 2005. EURASIP.

[9] H. Fletcher: "Mode locking in nonlinearly excited inharmonic musical oscillators," Vol. 64, pp. 1566-1569 *J. Acoust. Soc. Am.*, 1978.

[10] Judith C. Brown: "Frequency ratios of spectral components of musical sounds," *J. Acoust. Soc. Am*. vol. 99, no. 2, pp. 1210-1218, February 1996.

[11] B. Raj and P. Smaragdis: "Latent Variable Decomposition of Spectrogram for Single Channel Speaker Separation," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* , pp. 17-20, Oct. 2005.

[12] R. Dannenberg and C.Raphael: "Music Score Alignment and Computer Accompaniment," *Communications of the ACM, 49(8) (August 2006), pp. 38-43.*

[13] N. Ono, K. Miyamoto, J. Le Roux, H. Kameoka, and S. Sagayama: "Separation of a Monaural Audio Signal into Harmonic/Percussive Components by Complementary Diffusion on Spectrogram," *EUSIPCO.*, 2008

[14] Xavier Rodet: "Musical Sound Signal Analysis/Synthesis: Sinusoidal+Residual and Elementary Waveform Models," *IEEE Time-Frequency and Time-Scale Workshop 97*, Coventry, Grande Bretagne, 1997

[15] J Bouvrie and T Ezzat.: "An incremental algorithm for signal reconstruction from short-time fourier transform magnitude." *9th Intl. Conf. on Spoken Language Processing,* 2006.

[16] R. E. Kalman: "A New Approach to Linear Filtering and Prediction Problems," *Transaction of the ASME - Journal of Basic Engineering*, 35-45. March 1960.

[17] R.L. Eubank: *A Kalman Filter Primer*, Chapman & Hall/CRC, 2005.