Implementation of Transient and Tonal Molecular Pursuit using Discrete Wavelet Transform (DWT) and Modified Discrete Cosine Transform (MDCT) Atoms

MUMT 622 – 2018 Winter

Final Project

Harish Venkatesan

1. Introduction

Matching pursuit, as first proposed by Mallat and Zhang [1], is a greedy algorithm for sparse representation of signals as a weighted sum of atoms that belong to an over complete dictionary. A signal may be represented as follows

$$x = \sum_{i=0}^{N} \alpha_{\lambda_i} u_{\lambda_i} + R_N \tag{1}$$

where, α_{λ_i} is the weight of the atom, u_{λ_i} is an atom that belongs to a dictionary of atoms D, λ_i is the vector of parameters of a selected atom and R_N is the residual signal. The dictionary of atoms is so selected such that the signal to be decomposed has strong similarities with the atoms of the dictionary, thereby achieving accurate representation.

The basic principle of the matching pursuit algorithm involves several steps; first, the inner product of the given signal is computed with every atom of the dictionary, then the atom with the highest correlation with the signal is selected and removed from the signal. The inner products of all the atoms are then updated with the obtained residual. This process is continued until a stop condition is achieved. This stop condition may be the energy of the residual falling below a threshold or the available bit budget has been reached in compression applications.

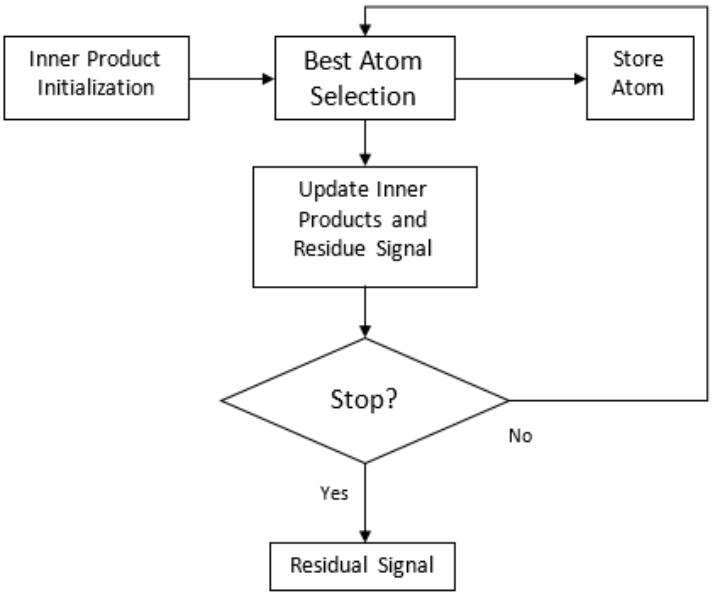

Figure 1: Matching Pursuit

However, the matching pursuit has computationally expensive steps. Searching for the highest correlated atom when the dictionary is large and updating inner products are two very costly steps. The former step may be made cost effective by sorting the atoms hierarchically or by using a 'weak matching pursuit' where search for an atom is stopped as soon as a near optimum atom has been found. The latter step can be made quicker by precomputing inner products to generate large look up tables.

The aim of this project is to implement a fast matching pursuit algorithm as proposed by Laurent Daudet in his paper entitled "*Sparse and Structured Decompositions of Signals With the Molecular Matching Pursuit*" [2], where at every step a set of m_i significant atoms that

form a molecule M_i are selected, thereby reducing the number of inner product updates to be performed. The proposed algorithm makes use of both time-frequency and time-scale atoms to select atoms based on the structural information. An added advantage of the utilizing the structural nature of signals is that these structures can be encoded much more efficiently than the individual samples.

The following section introduces the signal model and the dictionary of atoms used in the algorithm, section 3 talks about building the molecules, section 4 presents the molecular matching pursuit algorithm, section 5 explains how pre-echo artefact caused due overlap-and-add can be reduced, and finally, section 6 concludes with some closing remarks.

2. The Signal Model and Dictionary

In the paper by Daudet, audio signals are considered to consist of three part, namely tonal part (short lived components of certain frequency), transient part (attacks of notes) and residual part (stochastic signals such as breath) which are additive in nature, i.e., they can exist simultaneously. The dictionary of atoms chosen for this algorithm consists of two over-complete dictionaries, $D = C \cup W$, where the dictionary C consists of atoms of Modified Discrete Cosine Transform (MDCT) and models tonal part of the signal and the dictionary W consists of Discrete Wavelet Transform (DWT) atoms and models the transient part of the signal.

In the previously presented paper (Assignment 3 – Instrument-specific harmonic molecular matching pursuit [3]), the dictionary consists of chirped-gabor atoms and the molecules are built based on the harmonic nature of signals. The harmonic molecular matching pursuit algorithm would not work for non-harmonic signals such as percussive sounds and even slightly inharmonic sounds such as piano. The algorithm presented by Daudet in [2] would work for all signals that conform to the above mentioned model as it incorporates atoms that can model transient signals.

2.1 Modified Discrete Cosine Transform (MDCT) Atoms

Modified Discrete Cosine Transform can be considered similar to the short time fourier transform, except that the basis function does not consist of the imaginary part, i.e., the basis function of the MDCT is a modulated cosine. The MDCT coefficients of a signal are given by

$$\beta_{p,k} = \langle g_{p,k}, x \rangle \tag{2}$$

which are the inner product of the signal with lapped cosines at time instances p and frequency indices k. The basis function $g_{p,k}$ is given by

$$g_{p,k}[n] = g_p[n] \sqrt{\frac{2}{L}} cos\left[\frac{\pi}{L}\left(k + \frac{1}{2}\right)\left(n + n_p\right)\right] \tag{3}$$

where $g_p[n]$ is the window function, which is generally a sine function of L samples which is equal to half the time period of the sine function, with an overlap of L/2, hence providing perfect reconstruction and n_p is the time instant of analysis. This transform is particularly useful in compression algorithms as it results in the same number of coefficients as the number of samples of the time-domain signal, unlike the STFT.

As it can be seen from eq. (3), the MDCT is sensitive to phase shifts (time variant) since it does not retain phase information of the signal. Hence, MDCT coefficients do not represent the spectrum of the signal and the coefficients must be regularized in ordered to identify spectral peaks. In [4], Daudet proposes a regularization scheme which would help us in identify tonal molecules (discussed later).

Consider a sinusoidal signal of frequency ω and phase ∅. The MDCT coefficients of this signal would be given by

$$\beta_k = \frac{\sqrt{2L}}{2\pi} \frac{\sin(\pi f)}{(f-k)(f-k-1)} cos\left[\frac{\pi}{2}(k-k_o)+\psi\right] \quad (4)$$

where $f = \frac{\omega L}{\pi}$, $k_o = \lfloor f \rfloor$ and $\psi = \emptyset + \frac{\pi}{2}\left(k_o + \frac{1}{2}\right) + \frac{\pi(L-1)}{2L} f$ (complete derivation given in [4]).

The pseudo-spectrum of the signal is defined as

$$S_{p,k} = \left({\beta_{p,k}}^2 + (\beta_{p,k+1} - \beta_{p,k-1})^2\right)^{1/2}, \text{ for } k = 0,1,\dots,L-1 \quad (5)$$

It can be shown that the pseudo-spectrum would achieve maximum at $k = k_o$regardless of the phase of the signal, hence giving a good estimate of where in our time-frequency space the signal's tonal component lies (refer fig. 2). The proof for this is given in the appendix of [4].

Considering tonal signals are short-lived stationary signals, we can derive a measure known as the local tonality index given by

$$T_{p,k} = \frac{1}{W} \sum_{i=0}^{W-1} S_{p+i,k} \quad (6)$$

where W is known as the persistence constant, which corresponds to the duration in which the tonal signal is considered to be stationary. This helps in avoiding erroneous detection as tonal parts in a signal which could be due to noise. We will see in section 3.1 how a tonal molecule is extracted from the input signal.

2.2 Discrete Wavelet Transform (DWT) Atoms

Discrete Wavelet Transform performs well in identifying transients in a signal. Transient atoms can be represented using two parameters; the scaling coefficient b and the time coefficient a. A wavelet would be localized around time $t = 2^b a$, where $b = 0,1,\dots,J$, where J is the largest scale, which gives the lowest resolution approximation of the given signal. The DWT coefficients of the signal are obtained by

$$\alpha_{b,a} = \langle w_{b,a}, x \rangle \quad (7)$$

where $w_{b,a}$ is a wavelet function at a scale b. In the proposed algorithm it is suggested that $J = 8$ would be sufficient to detect transients in any given signal with a fair amount of accuracy. The wavelet function is chosen such that it is well localized in time. The suggested wavelets are the Haar wavelet and the Daubechies-4 wavelet due to their good time localization.

Every wavelet atom has two children at the immediate lower level. Hence, wavelet representation of a signal can be built in the form of a fully connected dyadic tree shown in grey in figure 4. Transient molecules are built by choosing those wavelet coefficients

that form a fully connected branch till the coefficient at the highest scale (discussed in more detail in section 3.2).

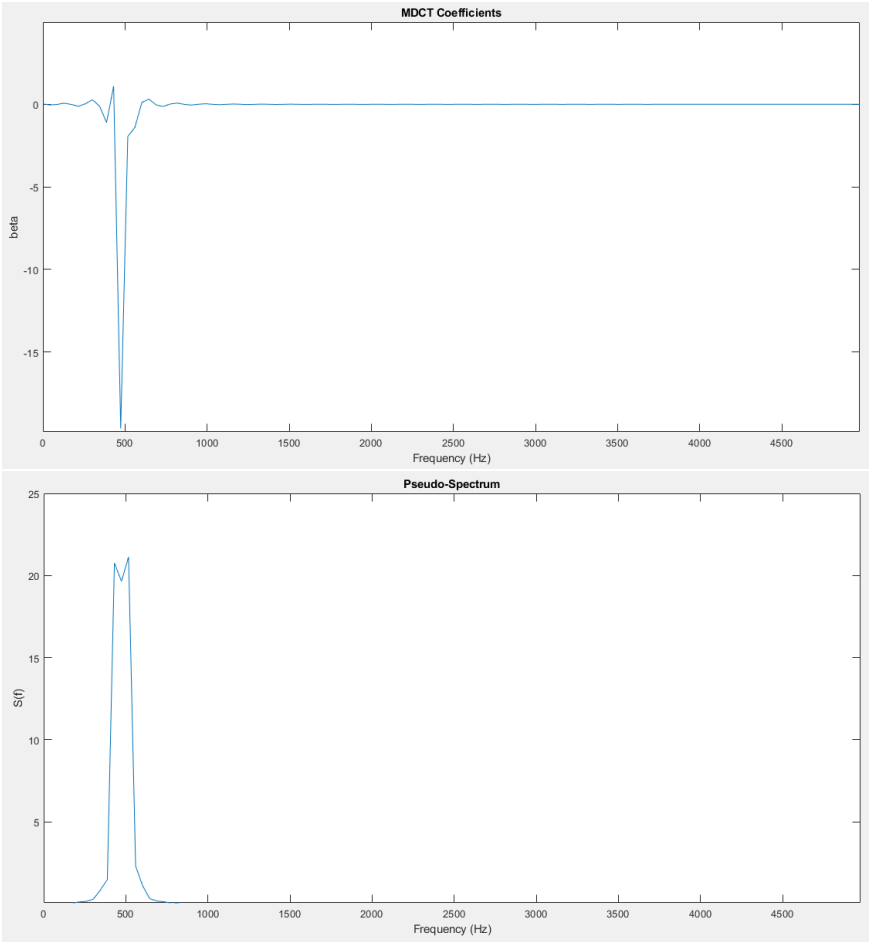

Figure 2:MDCT Coefficients and Pseudo-Spectrum of sine wave of 500Hz. Pseudo-spectrum shows clear peak around 500 Hz, whereas the MDCT coefficients do not indicate the frequency of the signal.

Transience in a signal is detected using a measure known as the regularity modulus κ [5]. This modulus takes into consideration all coefficients from the highest scale to the lowest scale and indicates the amount of energy spread across the spectrum. It is given by

$$\kappa[2t] = \frac{1}{J} \sum_{(b,a)\in B_t} \left|\alpha_{b,a}\right| \tag{8}$$

where B_t is the set of all ancestors of the smallest scale coefficient that form a full branch.

3. Building Molecules

3.1 Constructing Tonal Molecules

A tonal molecule is constructed around a given sample by picking out MDCT coefficients by a thresholding operation. Starting from a given pseudo-spectral coefficient at frequency k_0 and analysis frame p_0, we start to traverse forward until a the pseudo-spectral coefficient at an analysis frame $p + 1$ falls below a third of the coefficient at p, or it falls below an a priori threshold as given in eq. (9). This marks the end frame of molecule $p_{end} = p$.

$$S_{p+1,k} < \frac{S_{p,k}}{3} \; or \; S_{p+1,k} < \varepsilon_{stop} \tag{9}$$

The beginning of the molecule is identified by a similar operation traversing backwards starting from p_0. A tube of three MDCT coefficients centred at k_0 are picked out from the frames between p_{begin} and p_{end}, which would be sufficient to represent the amplitude, frequency and phase [6] of the stationary sinusoid. Then a post-thresholding operation is done in order to remove those MDCT coefficients whose absolute value falls below a threshold ε_{coeff}.

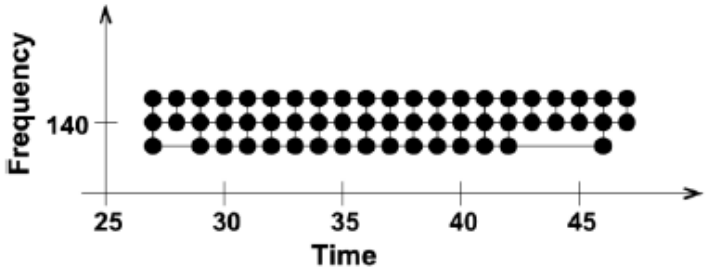

Figure 3: Selected MDCT coefficients after post-thresholding.

3.2 Constructing Transient Molecules

Once the time localization of a transient part of the signal is identified from the regularity modulus, the coefficients of all atoms of the DWT tree that it belongs to are selected. Then, starting from the smallest scale, those coefficients which fall below a threshold ε_{coeff} are removed, stopping at the coefficient on the branch which is above the threshold, ensuring that the tree remains fully connected. The following figure shows a selected transient DWT tree.

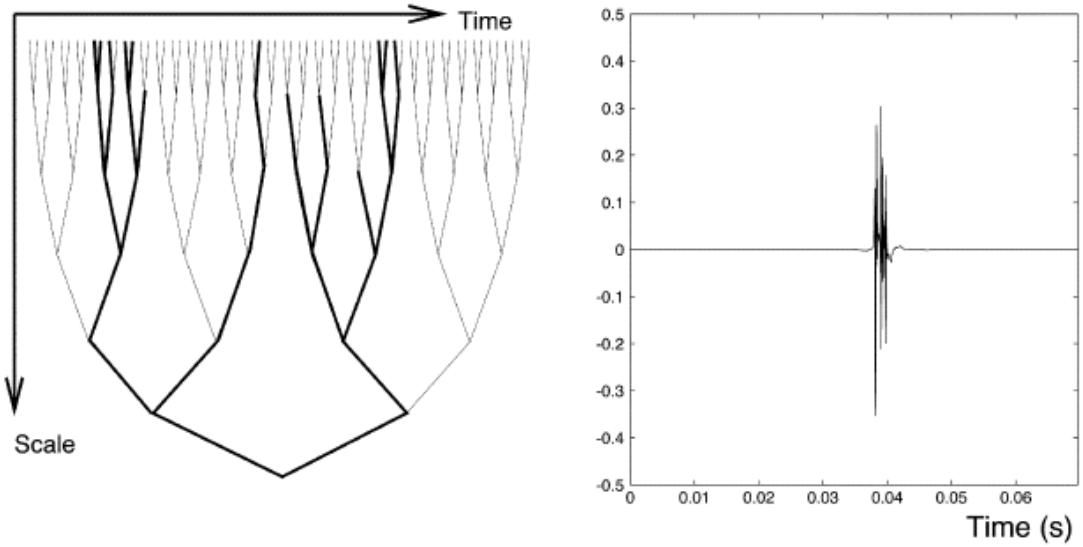

Figure 4: Selected DWT tree (shown in thick lines above), corresponding transient signal shown below.

4. The Molecular Matching Pursuit Algorithm

The below flow diagram (figure. 5) summarizes the proposed molecular matching pursuit algorithm.

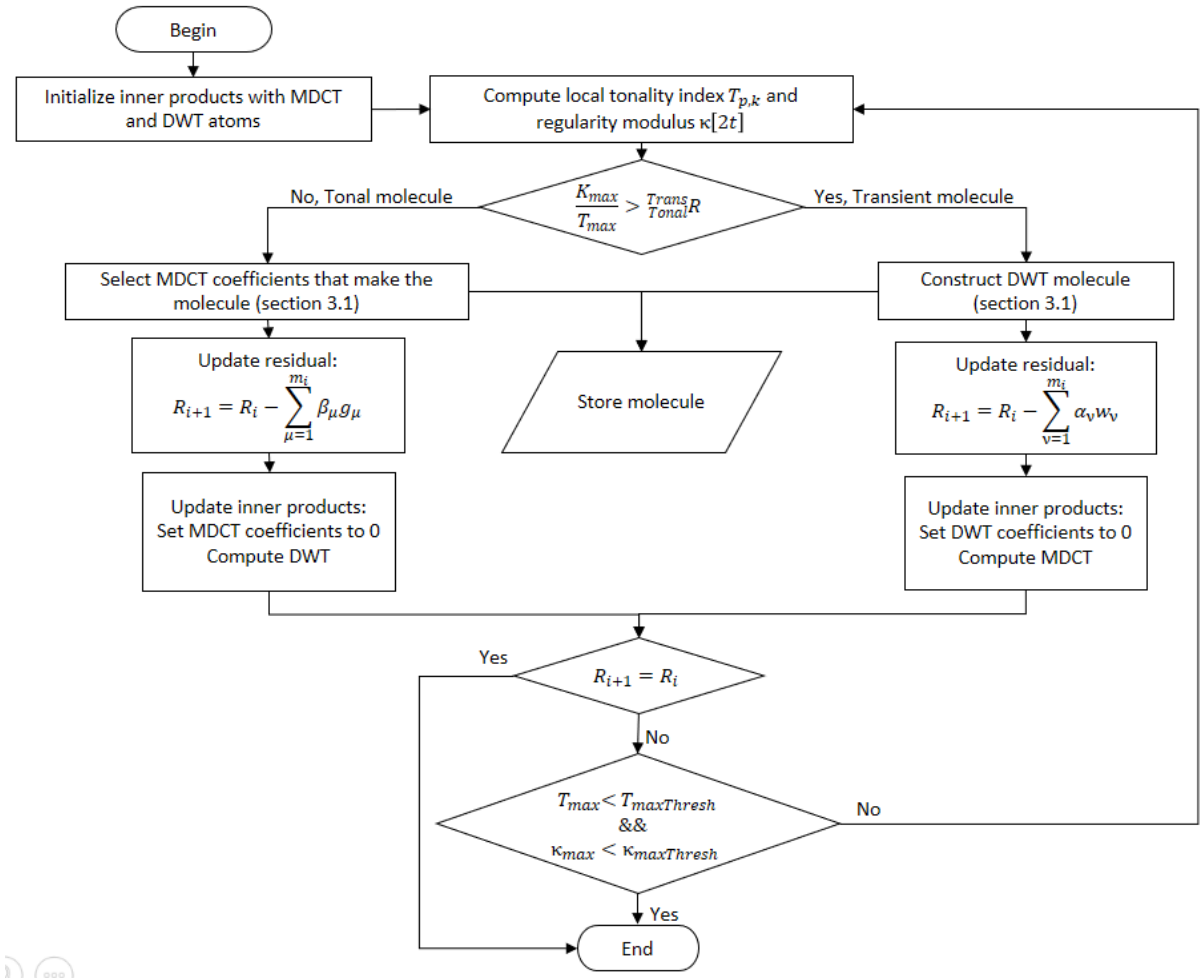

Figure 5: The molecular matching pursuit algorithm

Apart from reduction in computation cost by selecting a group of atoms at every iteration, the algorithm further reduces cost by computing inner products for only one set of atoms at every iteration. At any given iteration, if the chosen molecule is tonal, the MDCT coefficients that correspond to those atoms are set to 0 and only the DWT coefficients are recomputed. Similarly, if the chosen molecule is a transient, the DWT coefficients that correspond to those atoms are set to 0 and only the MDCT coefficients are recomputed.

The proposed algorithm requires two more thresholds, one for the maximum local tonality and one for the maximum regularity index, below which if the computed maximum indices fall, the algorithm is stopped. This makes it necessary for one to tune several parameters for different signals in order to achieve a good performance. It would be more convenient if there was another way to stop the computation, perhaps the energy of the residual signal. In our implementation, an additional condition is enforced; if the residual signal does not change in consecutive iterations, the algorithm would stop.

5. Application: Pre-echo Suppression

Due to the overlap-and-add technique used in the MDCT analysis of the signal with long windows, the attacks of notes get smeared and leads to a very audible artefact. Since the molecular matching pursuit algorithm proposed by Daudet separates signals into their

transient and tonal parts, the tonal part alone can be processed in order to suppress this artefact.

The pre-echo suppression is done by comparing a detected tonal molecule of 3-bin width with another molecule which is of a wider band around the same centre frequency. Accurate estimation of attack time is made as given in eq. 10.

$$\tau_i = \arg\min_t \left\| \sum_{\mu \in \nu_i} \beta_\mu g_\mu - H_t \sum_{\mu \in T_m} \beta_\mu g_\mu \right\|_2 \tag{10}$$

Then, the signal is truncated before the estimated time of attack τ_i using the Heavyside Step function H_{τ_i} . Figure 6 shows the working of the pre-echo suppression.

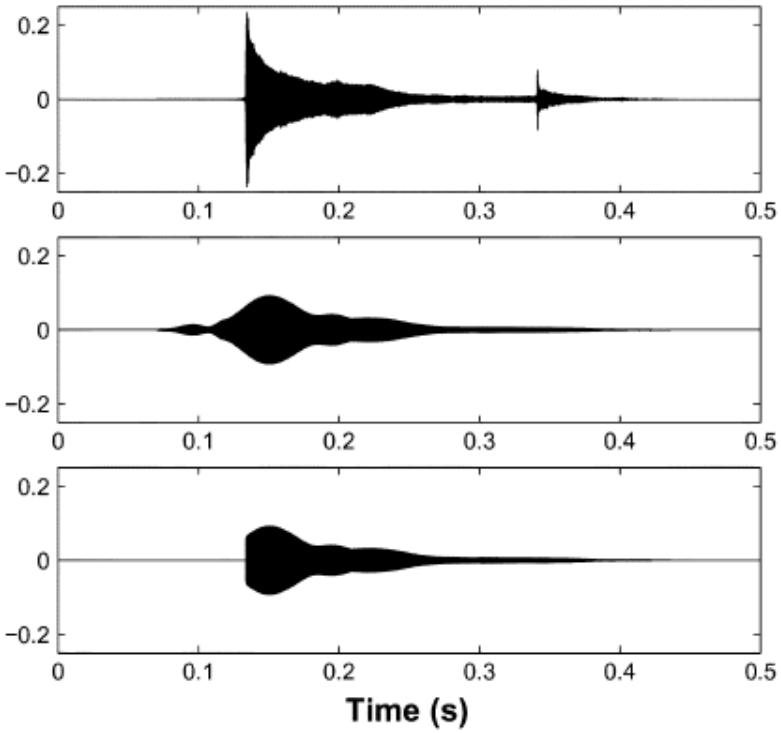

Figure 6: Pre-echo suppression. WIdeband molecule (top), selected molecule (middle) and truncated molecule (below). (Source [2])

6. Conclusion

The molecular matching pursuit proposed by Daudet in [2] was successfully implemented on Matlab with satisfactory performance. However, a few changes were made to the algorithm in order to achieve required performance. Firstly, an additional stop condition was added: if the residual signal does not change in two successive iterations, the algorithm would stop. Second, the regularity modulus for the transient molecule detection was implemented as given in [5], by adding an additional parameter γ to make the algorithm selective to transients with greater energy at the higher end of the spectrum, as given by eq. 11.

$$\kappa[2t] = \frac{1}{J} \sum_{(b,a) \in B_t} \left| \alpha_{b,a} \right|^\gamma, \gamma \leq 1 \tag{11}$$

Finally, an additional weighting parameter ${}^{Trans}_{Tonal}R$ was included in order to give more preference to one type of molecule over another at any given iteration. The decision whether the most prominent molecule is tonal or transient is slightly modified as follows.

$$If\ \frac{K_{max}}{T_{max}} > {}^{Trans}_{Tonal}R, Detected\ molecule = Transient$$
$$Else, Detected\ molecule = Tonal$$

Figure 7 shows a snapshot of the program developed in this project. The figure shows how well the algorithm performs for a given set of tuned parameters on a recorded guitar melody. We can see that the transients are detected very well by the algorithm but the tonal part is not satisfactorily extracted, leaving behind a large residual. Since the implementation of the algorithm does not take into consideration the time varying nature of the frequencies in the signal, a considerable amount of tonal information is left behind during tonal molecule detection and extraction. This algorithm could be very easily extended to incorporate the time-varying nature of frequencies and amplitudes in a signal.

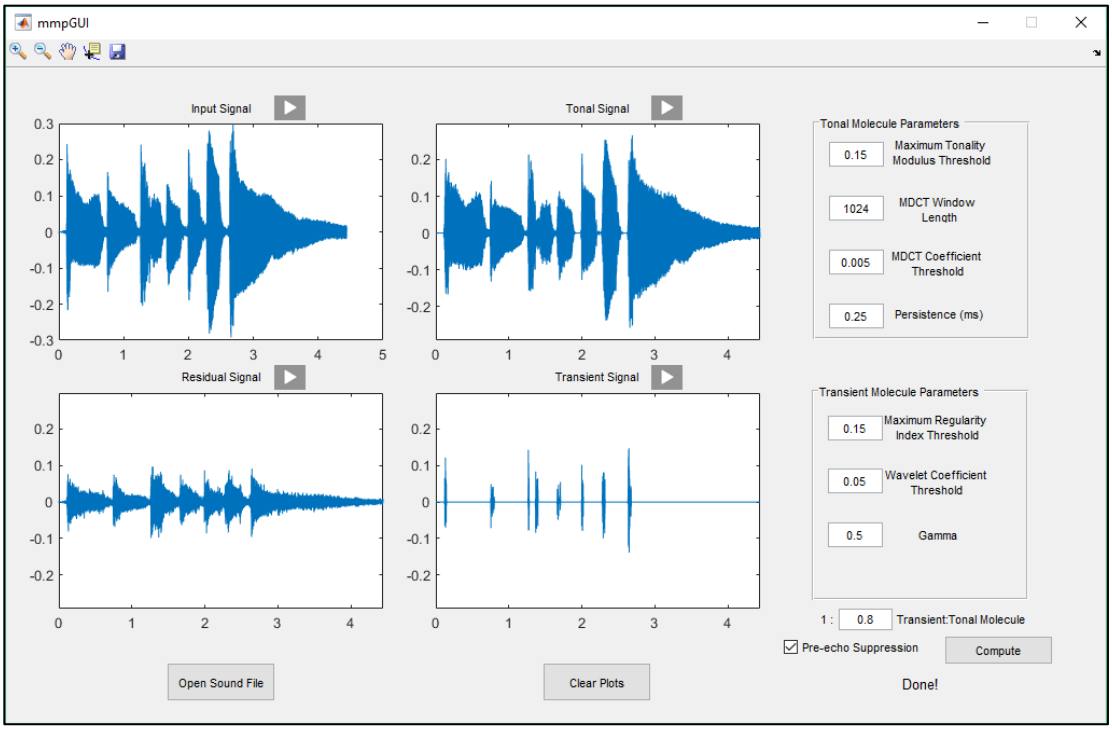

Figure 7: Molecular Matching Pursuit applied on a recorded guitar track.

7. References

[1] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[2] L. Daudet, "Sparse and structured decompositions of signals with the molecular matching pursuit," in *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 5, pp. 1808-1816, Sept. 2006. doi: 10.1109/TSA.2005.858540

[3] P. Leveau, E. Vincent, G. Richard and L. Daudet, "Instrument-Specific Harmonic Atoms for Mid-Level Music Representation," in *IEEE Transactions on Audio, Speech, and*

Language Processing, vol. 16, no. 1, pp. 116-128, Jan. 2008. doi: 10.1109/TASL.2007.910786

[4] L. Daudet and M. Sandler, "MDCT analysis of sinusoids: exact results and applications to coding artifacts reduction," in *IEEE Transactions on Speech and Audio Processing*, vol. 12, no. 3, pp. 302-312, May 2004. doi: 10.1109/TSA.2004.825669

[5] Daudet, Laurent, Stéphane Molla, and Bruno Torrésani. "Transient detection and encoding using wavelet coefficient trees." *18° Colloque sur le traitement du signal et des images, FRA, 2001*. GRETSI, Groupe d'Etudes du Traitement du Signal et des Images, 2001.

[6] S. Merdjani and L. Daudet, "Direct estimation of frequency from MDCT-encoded files," in *Proc. DAFx Digital Audio Effects Workshop*, London, U.K., 2003.