Extended Model Variety Analysis for Integrated Processing and Understanding of Signals *

E. Dorken, S.H. Nawab
ECS Department
Boston University

V.R. Lesser
Computer Science Department
University of Massachusetts

Abstract

In this paper, we extend our previous work on model variety analysis *of a signal processing algorithm with respect to the class of all input signals that may potentially arise in a given signal understanding application. This analysis has two related objectives. The first objective is to partition the set of all possible signals in the application domain into two sets according to whether each signal is correctly or incorrectly processed by the signal processing algorithm under consideration. The second objective of model variety analysis is to characterize the nature of the distortions in the signal processing output for the cases where the input signal is incorrectly processed. The results of model variety analysis are useful for designing signal understanding systems for applications where it is necessary for the signal processing to be carried out in a situation-dependent manner. Model variety analysis and its usefulness for the design of signal understanding systems are illustrated in this paper through examples involving the use of STFT processing for a sound understanding application.*

1 Introduction

The analysis of signal processing algorithms has traditionally been motivated by considerations such as computational efficiency, finite precision effects, and performance degradation when the signals for which an algorithm is designed are contaminated by noise. The need for a new type of analysis of signal processing algorithms has been pointed out [1] in the context of signal understanding systems that must carry out situation-dependent signal processing. This type of analysis of a signal processing algorithm has been referred to as *model variety analysis*. The basic idea behind such analysis is to determine how a given signal processing algorithm performs on a class of signals which includes signals which are not compatible with the assumptions under which the signal processing algorithm was originally designed. The results of such an analysis may be used by a signal understanding system for one of the two following purposes:

- to decide whether or not to use the given algorithm in a situation where the input signal is likely to belong to a particular class of signals.
- to decide whether the output from a particular application of the given algorithm may have been due to an input signal for which the use of that algorithm is inappropriate.

We have previously reported [1, 2] a relatively simple example of model variety analysis of the short-time Fourier transform (STFT) algorithm. We have now extended the scope of that example and in the process we have attempted to establish a methodology for carrying out model variety analysis in general. The primary purpose of this paper is to present the results of our extensions to the model variety analysis of the STFT algorithm and to indicate the generalizable aspects of the methodology used in the example.

In section 2 we provide some background on why situation-dependent signal processing may be required in some signal understanding applications. In particular, we use the example of a sound understanding application for which we have developed an experimental knowledge-based signal understanding system [3, 4]. That system performs STFT signal processing and utilizes the results of the model variety analysis presented in this paper to guide the situation dependent application of STFT algorithms with different analysis-window lengths, fft lengths, and temporal-decimation factors. In section 3, we describe our latest results on model variety analysis of the STFT algorithm.

2 Background

In this section we discuss why situation-dependent signal processing is deemed necessary for some signal understanding applications. In particular, we point out that the signal processing requirements for a particular situation depend upon the particular mix of transient, steady-state, and noise-like characteristics in the signal data corresponding to the situation. Consequently, for an application in which the mix of such characteristics varies considerably from situation to situation, it is crucial for a signal understanding system to be able to adjust its signal processing in a situation-dependent manner.

Imagine a household robot that has a hearing capability. This robot can tell when a smoke alarm goes

*This research was sponsored in part by the Rome Air Development Center (RADC) of the Air Force Systems Command under contract number F30602-91-C-0038, in part by the Office of Naval Research under a University Research Initiative grant, ONR N00014-86-K-0764, and in part by NSF under contract number CDA 8922572

off, when a baby is crying, or when the telephone is ringing etc. and can take appropriate action in response to such events. Alternatively, imagine an assistive device for the hearing impaired that can alert a deaf person about important events such as ringing fire alarms, door knocks, the sound of an oven buzzer etc. Finally, consider a computer that can understand verbal instructions in an office environment that may include other sounds such as ringing telephones, the noise of a copy machine, music etc. A generic activity involved in all these tasks is *sound understanding*. Broadly speaking, a sound understanding system must be able to obtain evidence from sound signals in order to produce descriptive assertions about the events taking place in a sound producing environment. In a realistic environment there is typically an enormous variety of possible sound producing events. Furthermore, these events can occur simultaneously and with different relative loudnesses and durations. Consequently, the characteristics of the signals received by a sound understanding system also tend to vary enormously. It is generally considered unrealistic to design a single signal processing algorithm that can perform satisfactorily under such a diverse set of conditions. The more realistic approach is for a sound understanding system to have access to a database of different signal processing algorithms from which it can intelligently select the most appropriate algorithms to use in a situation-dependent manner.

In our experimental sound understanding system, the database of signal processing algorithms consists of STFT algorithms with different values for the STFT control parameters (window-length, fft-length, and temporal decimation factor). Also considered part of the STFT processing is a signal detection procedure that picks the N most prominent spectral peaks of the STFT for each analysis window position. The parameter N is considered yet another control parameter of the STFT processing. Selection of a particular STFT algorithm from the database is equivalent to selecting the values for each of the control parameters.

A concrete illustration of the need to adjust the STFT control parameters in a situation-dependent manner can be obtained by considering two simple sound producing events: a car passing by and a hairdryer being turned on. The first event generates a prominent harmonic signal with several harmonics with an interharmonic spacing of about 90 Hz. The second event generates a prominent narrowband signal with a strong frequency modulation. The processing results for the two events with STFT's that utilize a 2048 point analysis window and a 256 point analysis window are illustrated in Figure 1. From the figures, it is clear that the window length which is appropriate for one event is inappropriate for the other.

If we now imagine these two events taking place simultaneously, using either one of the window lengths for the STFT would lead to incorrect results. In such situations a compromise window length would be needed. In some situations, the uncertainty principle intervenes to make it impossible to correctly process the data with one particular window length. For such situations, our sound understanding system has

Figure 1: STFT processing output for two sound producing events. A) "hairdryer turned on", window length is 2048. B) "hairdryer turned on, window length is 256. C) "car passing by", window length is 256. D) "car passing by", window length is 2048.

the capability to first process the signal data with one window-length and then to *reprocess* the signal data with another window length. A thorough discussion of how this can be practically accomplished is given in [1].

3 Model Variety Analysis

In this section we describe our latest results on model variety analysis of the STFT algorithm. These results were obtained by extending the analysis reported in [2]. There are two basic objectives in performing model variety analysis:

- To partition the class of possible input signals into two sets according to whether each signal is correctly or incorrectly processed by the signal processing algorithm under consideration.
- To characterize the nature of the distortions in the output of the signal processing algorithm for the cases where the input signal is incorrectly processed.

How model variety analysis may be performed to achieve each of the above two objectives is described in sections 3.1 and 3.2 respectively.

3.1 Partitioning

The algorithm under consideration is the STFT. It is assumed that the STFT has window-length N_w, fft-length N_{fft}, and temporal decimation L. The class of possible input signals is considered to include linear combinations of multiple finite-duration sinusoids. The individual sinusoids may be linearly modulated in frequency. More precisely, each signal $x(n)$ in this class may be expressed in the following form:

$$x(n) = \sum_{k=1}^{N} A_k cos[(\omega_k^o + \frac{1}{2}\alpha_k n)n][u(n-n_k^s) - u(n-n_k^e)] \quad (1)$$

where,
A_k is the amplitude,
ω_k^o is the frequency offset,
α_k is the frequency modulation slope,
n_k^s is the start time,
n_k^e is the end time,
$\omega_k(n) = (\omega_k^o + \alpha_k n)$ is the instantenous frequency,
$d_k = (n_k^e - n_k^s)$ is the duration of the k^{th} sinusoid.

The various parameters of each $x(n)$ are termed the signal generation parameters for that $x(n)$. We also assume that the signal generation parameters of the class of input signals are such that it is possible to distinguish between the different signals if estimates of the signal generation parameters are obtained in accordance with the following "compatibility criteria".

compatibility criteria: The criteria for compatibility between the actual values of the signal generation parameters $(\omega_k(n), d_k)$ and the estimated values $(\hat{\omega}_k(n), \hat{d}_k)$ are given as:

$$\left| \frac{\hat{\omega}_k(n) - \omega_k(n)}{\omega_k(n)} \right| < 0.1 \quad \text{(A)}$$

$$\left| \frac{\hat{d}_k - dk(n)}{d_k(n)} \right| < 0.1 \quad \text{(B)}$$

$$||\hat{\omega}_k(n) - \hat{\omega}_k(m)| - |\omega_k(n) - \omega_k(m)|| < \frac{2\pi}{1000} \quad \text{(C)}$$

Criteria A and B stipulate the requirement that the estimate of instantenous frequency and the duration of sinusoids be within 10% of their actual values. Criterion C stipulates the requirement that the frequency changes of $\frac{2\pi}{1000}$ radians or higher must be detected.

Following the procedure detailed in [2], we then derive the *data-model* for the STFT algorithm with respect to the class of input signals described above. The data-model specifies a set of conditions which the signal generation parameters of any input signal must satisfy in order for the signal to be correctly processed by the STFT. In this analysis, a signal is considered correctly processed if the significant peaks detected in the STFT output are compatible with the signal generation parameters of that signal. Here compatibility is determined in accordance with the compatibility criteria described above. The data-model we have derived for the STFT with respect to the class of input signals under consideration is presented in Table 1. The conditions in this data-model specify how the signal generation parameters of a signal must be related to the control parameters of the STFT in order to be correctly processed by that STFT. The data-model therefore represents a partitioning of the set of possible input signals into two subsets according to whether or not they are correctly processed.

1.	$min(\omega_k(n))_{k=1..N} > \frac{\pi}{N_w}$
2.	$0.1min(d_k)_{k=1..N} > L$
3.	$(n_m^s - n_k^e)_{m,k=1..N,m\neq k} > N_w$
4.	$max(\alpha_k)_{k=1..N} < \frac{2\pi}{1000N_w}$
5.	$\lvert\omega_m - \omega_k\rvert_{m,k=1..N,m\neq k} > \frac{\pi}{N_w}\left[\frac{A_m}{A_k} + 1\right]$ and $A_m > A_k$
6.	$0.1min(\omega_k)_{k=1..N} < \frac{2\pi}{N_{fft}}$
7.	$\lvert\omega_m - \omega_k\rvert_{m,k=1..N,m\neq k} > \frac{2\pi}{N_{fft}}$
8.	$0.1min(d_k)_{k=1..N} > N_w$
9.	$max(\alpha_k)_{k=1..N} < \frac{2\pi}{1000L}$

Table 1: STFT data-model.

3.2 Characterization

We now consider the problem of characterizing the results of incorrect processing when the conditions in the data-model are violated. It should be noted that the data-model of section 3.1 was derived with repect to a larger class of possible input signals than those considered for the data-models in [1] and [2]. The reason for this is that in our previous research we were primarily concerned with showing that the set of signals that would be incorrectly processed is non-empty. In contrast, the research reported in this paper is also concerned with characterizing how the signal processing output violates the compatibility criteria when the input signal is incorrectly processed. To be useful in the design of a signal understanding systems, such a characterization must encompass as many possible input signals in the application as can be accomodated. Although a large number of possible input signals in the sound understanding application can be modelled by equation 1, it is clear that not all of them can be modelled this way. Our on-going research is aimed at expanding the class of input signals for which data-models can be derived. This is indeed the challenge of research in the area of model variety analysis of the STFT algorithm.

The methodology we have adopted for characterizing the results of incorrect processing is to construct *distortion operators* that specify how the signal generation parameters of every input signal are mapped into STFT output parameters if the signal is incorrectly processed. The analysis is conducted by considering each condition in the data-model separately and using the theory behind STFT processing to determine the mapping from signal generation parameters to STFT output parameters. We have derived such distortion operators corresponding to each of the conditions in the data-model of Table 1. A description of a few of those operators is given below.

1. **insufficient-data distortion:** The associated data-model condition (1, Table 1) is derived on the basis that the frequency of a sinusoid must be higher than $\frac{\pi}{N_w}$ in order to eliminate the effects of the interaction between the negative and positive frequency components of that sinusoid. When this condition is violated spectral leakage between the negative and positive frequency components of a sinusoid gives rise to a local maximum at 0 frequency. The following distortion operator specifies how instantenous frequency estimates deviate from their actual values when this distortion arises.

$$\omega_k(n) \in \left[0, \frac{\pi}{N_w}\right] \rightarrow \hat{\omega}_k = 0 \qquad (2)$$

All signals that satisfy the condition on the left hand side of equation 2 give rise to identical instantenous frequency estimates.

2. **low-time-resolution distortion:** The associated data-model condition (3, Table 1) is derived on the basis that two signals with identical frequencies should be separated in time by a duration greater than the duration of the STFT window length in order for STFT to resolve them. If this condition is violated, duration estimates deviate from their actual values as described in the equation below.

$$(d_k, d_m) \rightarrow \hat{d}_i = d_k + d_m + (n_m^s - n_k^e) \qquad (3)$$

In the equation above, d_k, d_m are the durations of two distinct sinusoids, n_m^s is the starting time of the m^{th} sinusoid, n_k^e is the end time of the k^{th} sinusoid, and $\hat{d}_i$ is the duration estimate associted with this sinusoid pair. Notice that there are many sinusoid pairs with different durations and different start and end times that give rise to the same duration estimate.

3. **low-frequency-discrimination distortion:** The corresponding data-model condition (6, Table 1) is derived on the basis that the fft-length must satisfy a lower bound in order that the frequency of a sinusoid be estimated within a tolerance given in compatibility criterion A. If this condition is violated, instantenous frequency estimates deviate from their actual values as specified by the equation below.

$$\omega_k(n) \in [\frac{\pi(2m-1)}{N_{fft}}, \frac{\pi(2m+1)}{N_{fft}}] \rightarrow \hat{\omega}_k = \frac{2\pi m}{N_{fft}} \qquad (4)$$

In the equation above, m is an integer such that $m \in \left[0, \frac{N_{fft}}{2}\right]$. All sinusoids that have a frequency within the range given by equation 4 give rise to identical instantenous frequency estimates.

References

[1] S.H. Nawab and V. Lesser, "Integrated Processing and Understanding of Signals," *Symbolic and Knowledge-Based Signal Processing*, edited by A.V. Oppenheim and S.H. Nawab, Prentice Hall, Englewood Cliffs, NJ, 1991.

[2] A.S. Weiss, E. Dorken, S.H. Nawab, "Model Variety Analysis: Determining the Need for Knowledge-Based Signal Processing", *1991 Asilomar Conference On Signals, Systems, and Computers*, 1991.

[3] V.R. Lesser et. al., "Integrated Signal Processing and Signal Understanding",*Umass Tech. Report*, 91-34, December 1991.

[4] E. Dorken, S.H. Nawab, E. Milios, "Knowledge-Based Signal Processing Applications," in *Symbolic and Knowledge-Based Signal Processing*, edited by A. V. Oppenheim and S. H. Nawab, Prentice Hall, Englewood Cliffs, NJ, 1991.