Fusing Multiple Reprocessings of Signal Data

Frank Klassner Victor Lesser Hamid Nawab†

Computer Science Department
University of Massachusetts
Amherst, MA 01003
{klassner lesser}@cs.umass.edu

†ECS Department
Boston University
Boston, MA 02125
hamid@engc.bu.edu

ABSTRACT

In the analysis of signals from complex environments, often it is not possible to rely on a single set of signal processing algorithms (SPAs) to produce a set of data correlates that permit meaningful interpretation. In such situations, what is needed is the structured fusion of data from multiple applications of SPAs (reprocessings) with different parameter values. We present the *Integrated Processing and Understanding of Signals* (IPUS) architecture as a framework for structuring interaction between the search for SPAs appropriate to the environment and the search for interpretation models to explain the SPAs' output data correlates. In this paper we describe our use of IPUS to control the integration of output from multiple SPA applications in a system for acoustic signal interpretation of household sounds.

1. INTRODUCTION

Traditionally, perceptual systems have been designed according to the architecture of Figure 1, with a fixed set of front-end signal processing algorithm (SPA) instances [6, 11]. By "SPA instance" we mean a generic algorithm (e.g. an N-point FFT) instantiated with specific control parameter values. These instances are chosen after careful analysis of the environment determines which combinations of SPA instances can compute "adequate" correlates for the interpretation component of the system. Regardless of what fusion methods are used by the interpretation component, the implicit assumption is that the SPA instances supplying the correlates are appropriate to the environment being monitored.

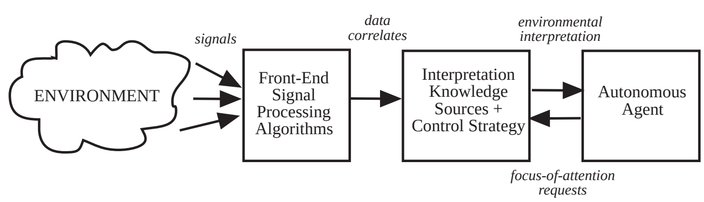

Figure 1: *Classic Signal Processing Architecture. Because the paradigm assumes that the front-end is tailored to the environment, often little or no formal feedback from the interpretation subsystem is incorporated.*

Consider the generic Short-Time Fourier Transform (STFT) algorithm [10] in the acoustic domain. An STFT SPA instance has particular values for its parameters, such as analysis window length, frequency-sampling rate, and decimation interval. Depending on a scenario's spectral features and their time-variant nature, these parameter values increase or decrease the instance's appropriateness for monitoring the scenario. An instance with a wide analysis window provides fine frequency resolution, and will be appropriate for generating correlates from scenarios containing sounds with time-invariant components. The same instance, however, does not provide fine time resolution, and will not be appropriate for generating correlates from scenarios containing sounds with time-varying components.

The traditional design paradigm's "appropriateness assumption" raises challenges for system designers when they attempt to apply it to complex environments. Complex environments have variable signal to noise ratios, unpredictable object behaviors, and many objects whose signatures can mask or otherwise distort each other. The environment can change dramatically from the state it was in when a perceptual system was designed and deployed. For example, speech recognition systems configured in a closed room may work well until someone opens a window and background traffic sounds lead to unacceptable performance.

Use of the traditional paradigm in complex environments often leads to front-end SPA sets that grow combinatorially with the possible object combinations and environmental states [5]. Complex environments do not lend themselves to comprehensive analysis for determining the appropriate front-end SPA instances. Thus, in an attempt to simplify design under the traditional paradigm, front-end designers often avoid highly specialized (sometimes low-cost) SPAs in favor of generalized SPAs that are immune to environmental shifts but whose correlates have less identification power.

In an effort to solve these problems from a knowledge-based perspective, we have developed the *Integrated Processing and Understanding of Signals* (IPUS) architecture [7, 8]. The architecture design makes no assumptions about the appropriateness of its front-end to the environment. Instead, it incorporates explicit knowledge about the signal processing theory underlying the set of front-end generic SPAs, and uses this information to decide when SPA instances are no longer appropriate, or need to have their correlates augmented with selective application of instances of specialized SPAs. With its elimination of the appropriateness assumption, the IPUS paradigm transforms the process of perception into two interleaved search processes: a search for plausible interpretations of SPA correlates and a search for appropriate SPA instances.

While performing these searches, the architecture engages in data reprocessing. We use this term to refer both to re-application of generic SPAs with different control parameters and to selective application of SPA instances to augment initial processings' deficiencies with respect to the environment [1]. The results of these reprocessings must be carefully maintained and fused, since they play an important role in constraining the framework's two search processes. During the fusion of results from several reprocessings, correlates obtained from different instances of the same or different generic SPAs applied to the same signal must be checked not only for consistency with the current interpretation of the environment, but also for consistency with each other. A reprocessing architecture must provide mechanisms for applying signal processing theory to distinguish between those discrepancies that arise due to inappropriate SPA usage and those that arise due to incorrect interpretations.

Using SPA instances as analogues to traditional hardware sensors, this paper examines the fusion issues in a reprocessing architecture and describes the mechanisms employed in our IPUS testbed. Although we will refer to examples from the acoustic interpretation testbed in which we initially developed IPUS, we believe that the issues in the paper are actually generic to all sensory modalities.

2. ARCHITECTURE BACKGROUND

This section provides a summary of IPUS. It is intended to serve as minimal background for appreciating the reprocessing architecture's fusion issues discussed in later sections. For more detail on the architecture, we refer the reader to [7] and [8]. For more detail on control within the IPUS framework, we refer the reader to [2] and [3].

2.1. Architecture Terminology

We define the *data-model* of an SPA instance with respect to a set of environment signals $\mathcal{S}$ as the set of conditions which various features of the members of $\mathcal{S}$ must satisfy in order to ensure that the instance produces correlates from which those signal features can be estimated adequately. When these conditions are satisfied by an environment's signal, the SPA instance is considered *appropriate* to the environment, and its correlates are said to be *undistorted.* If the data-model conditions are not satisfied by the environment's signal, the SPA instance is considered *inappropriate* to the environment and its correlates are said to be *distorted.*

In the following discussion, note that there is a difference between "discrepancies" and "distortions." A *dis-*

crepancy is a difference observed between an expectation for feature values in a signal and the actual values of the correlates extracted from the signal. *Distortions* are processes defined by formal signal processing theory that occur when an SPA instance is inappropriately applied to an environment's signal. Distortion processes are used to explain discrepancies. It is also possible for several distortion processes to explain the same discrepancies.

2.2. Reprocessing Loop

The generic IPUS architecture, with its primary data and control flow, appears in Figure 2a. Figure 2b shows its instantiation in the acoustic interpretation testbed. Two types of signal interpretation hypotheses are stored on the hierarchical blackboard: interpretations of correlates from current and past signal analyses, and expectations about the interpretations of data correlates from future analyses.

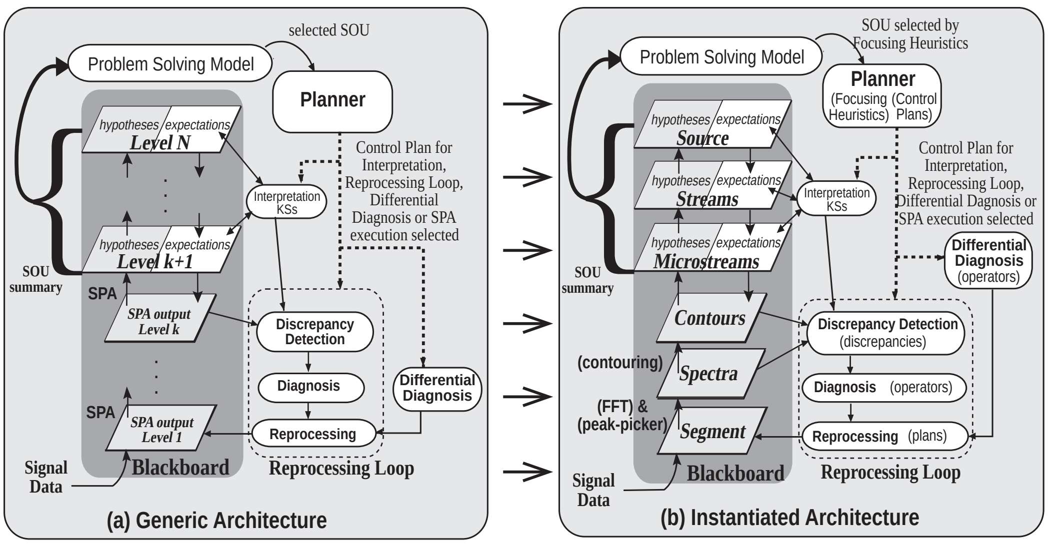

Figure 2: *Figure 2a shows the generic IPUS architecture, figure 2b shows the architecture instantiated for the sound understanding testbed. Solid arrow lines indicate dataflow relations. Dotted arrow lines indicate classes of plans that the planner can pursue when trying to reduce or eliminate particular uncertainties (discrepancies) in the problem solving model that were selected by the focusing heuristics. Knowledge added to the planner or system knowledge sources to instantiate the architecture for an application is shown in parentheses. The acronym "SOU" stands for Source of Uncertainty. SOUs are structures that represent factors (e.g. missing evidence or competing alternative interpretations) that affect interpretations' confidence levels.*

Our design of the IPUS framework assumes that signal data is submitted for analysis a block at a time. IPUS uses an iterative process for converging to the appropriate SPAs and interpretations. For each block of data, the loop starts by processing the signal with an initial configuration of SPAs. These SPAs are selected not only to identify and track the signals most likely to occur in the environment, but also to provide indications of when less likely or unknown signals have occurred. In the next part of the loop, a *discrepancy detection* process tests for discrepancies between the correlates of each SPA in the current configuration and (1) the correlates of other SPAs in the configuration, (2) application-domain constraints, and (3) the correlates' anticipated form based on high-level expectations. Architectural control permits this process to execute both after SPA output is generated and after interpretation problem solving hypotheses are generated. If discrepancies are detected, a *diagnosis* process then attempts to explain them by mapping them to a sequence of qualitative distortion hypotheses. These distortions are defined by formal signal processing theory such as Fourier analysis. The loop ends with a *signal reprocessing* stage that proposes and executes a search plan to find a new front-end (i.e., a set of instantiated SPAs) to eliminate or reduce the hypothesized distortions. After the loop's completion, if there are any similarly-rated competing top-level interpretations, a *differential diagnosis* process selects and executes a reprocessing plan to find correlates for features that will discriminate among the alternatives.

Although the architecture requires the initial processing of data one block at a time, the loop's diagnosis, reprocessing, and differential diagnosis components are not restricted to examining only the current block's processing results. If the current block's processing results imply the possibility that earlier blocks were misinterpreted or inappropriately reprocessed, those components can be applied to the earlier blocks as well as the current blocks.

Each time the data is reprocessed, whether for disambiguation of competing interpretations or for elimination of distortions, a new state in the SPA search space is tested for how well it eliminates or reduces distortions. This distortion elimination test is based on the assumption that the system's current state in the interpretation search space matches the actual context being observed. Failure during reprocessing to remove a hypothesized distortion after a bounded search in the SPA instance space leads to a new search in the interpretation space. This happens based on the following reasoning. The diagnosis explanation and reprocessing results represent an attempt to justify the assumption that the current interpretation is correct. If the diagnosis component cannot produce a theoretically plausible explanation for discrepancies or if the reprocessing component fails to remove discrepancies, there is a strong likelihood that the current interpretation is not correct and a new search is required in the interpretation space.

3. SENSOR FUSION IN A REPROCESSING ARCHITECTURE

In this section we discuss the fusion issues that arise in a reprocessing architecture. Specifically, we first discuss the fusion of observations of the same data made with different instances of the same generic SPAs, then we discuss the fusion of observations of the same data made with instances of different generic SPAs.

The *Model Variety Problem* [5, 9] focuses on the relationship between SPAs and the classes of signals for which they can produce undistorted correlates. A signal interpretation system needs to apply more than one instance of an SPA if there does not exist a single instance of the SPA that is appropriate to all possible environment signals. In other words, the input signal must always satisfy the conditions in the data-model for that SPA instance. A need for more than one SPA instance translates to the need for a variety of data-models. For such applications, we can therefore say that the given generic SPA has a model variety problem with respect to the environment. We use the term *Model Synthesis* to refer to the process of fusing interpretations and correlates obtained from different SPA instances applied to the same data in order to handle model variety problems. There are two classes of synthesis: refinement and integration.

Refinement occurs in the case when it is possible to find an SPA instance that removes all of the discrepancies noted in the correlates of an originally-applied instance of the same generic SPA. Reprocessing with this new instance produces correlates that only *refine* those of earlier processings. For example, consider an STFT SPA with 1024-point FFT-lengths applied to 256-point analysis windows (assume a 10KHz sample rate). Assume that two previously-observed frequency tracks drift toward each other to a separation smaller than 40 Hz. The resulting poor frequency resolution distortion would cause a discrepancy because expectations for two tracks would be violated. Only one track is observed, and at an unexpected frequency (see Figure 3). The correlates produced by reprocessing with an STFT instance having the same FFT-length but a 512-point analysis window would provide greater frequency resolution and would represent a refinement of the first STFT's correlates. In IPUS, correlate fusion in this case consists of replacing the original SPA's correlates in the blackboard with the second SPA's correlates. Interpretations based on these correlates simply have their constraints refined with the more detailed information.

Integration occurs when no single set of values for the control parameters of a generic SPA will eliminate all distortions in the SPA output at the same time. In this case we say that the generic SPA has an inherent model-variety problem with respect to the environment [9]. For example, in the acoustic domain no SPA can simultaneously provide infinite frequency and infinite temporal resolution. In these situations it is necessary for the reprocessing KS to integrate the results of several reprocessings, each of which removes only *some* of the observed distortions, and may introduce new distortions that must be ignored.

Consider the processing of an electric motor sound that contains a speed-change transition. Let us assume that

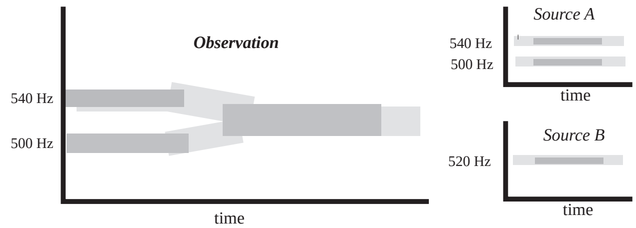

Figure 3: *A situation where refinement synthesis is possible. Darker shading indicates higher energy. When source A's tracks "drifted" toward each other, the FFT frequency resolving capability is exceeded and they appear merged as one track which could represent a poorly processed source A, the appearance of source B and the disappearance of source A, or the simultaneous presence of source A and source B.*

there have been no earlier speed-changes and that a front-end generic STFT has therefore had its control parameters set to values that ensure detection of two steady-frequency tracks. Let us further assume that the system has no expectation for a speed change at the current time; instead it expects the two steady-frequency tracks to continue as previously observed. However, when the portion of the signal containing the speed change is processed, the correlates shown in Figure 4 are produced.

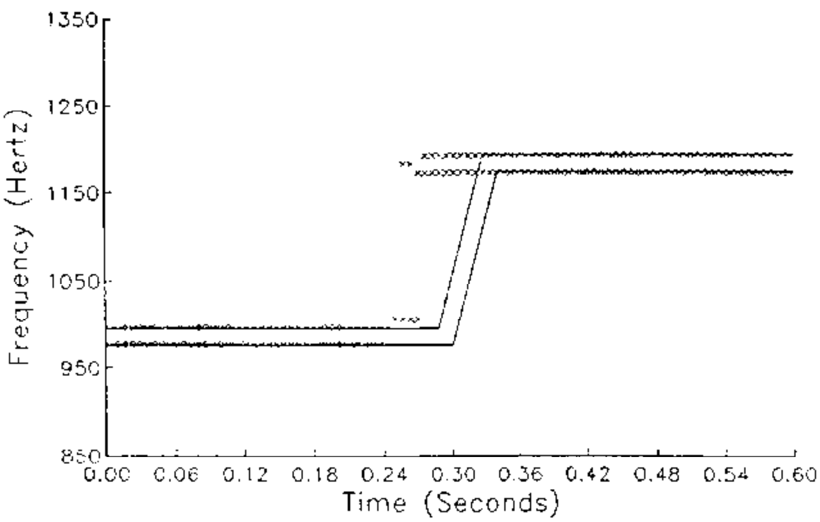

Figure 4: *STFT output of a motor sound as a situation where integration synthesis is desirable. Solid lines represent the actual frequency tracks while ×'s represent the STFT correlates (peaks). Poor time-resolution causes the motor speed-change interval to be undetected.*

The signal processing output is in conflict with the expectation of two steady-frequency components. The IPUS diagnosis component hypothesizes that the upper two tracks in Figure 4 are connected with the lower two tracks and that this connection is missing in the SPA output due to a time-resolution distortion. This distortion arises when the STFT-instance has its window-length parameter set to a relatively large value. Given this explanation, the IPUS reprocessing component selects a reprocessing plan that uses an STFT instance with a shorter analysis-window.

The subsequent execution of the reprocessing plan results in the correlates shown in Figure 5. As expected, evidence is obtained for the speed-change, but now poor frequency resolution resulting from the shorter window length does not resolve the two constituent frequency components. If IPUS were to register this as a new discrepancy, the system would become trapped in a discrepancy loop, since there is no single STFT instance that can capture both aspects of the signal. Thus, the system must be capable of anticipating the new distortion and instead *integrate* the data from Figures 4 and 5 as jointly representing evidence for the interpretation-model represented

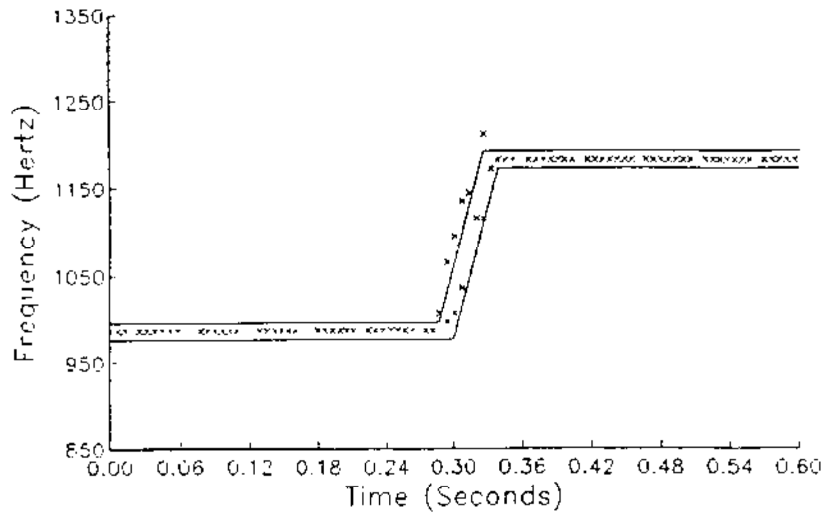

Figure 5: *Reprocessing correlates from applying an STFT with a shorter analysis window to the signal observed in Figure 4. Adequate time-resolution detects the motor change, but poor frequency-resolution causes the two tracks to merge into one.*

by the solid-lines.

In the above discussion, all of our examples have centered on reprocessing triggered by discrepancies detected between expectations and SPA correlates. However, it is often the case that an SPA instance's appropriateness to an environment (i.e. the absence of distortions in its correlates) must be determined by comparing its correlates with correlates from an instance of a different generic SPA. When discrepancies are detected, a system must be able to use them to find new SPA instances that can produce consistent correlates (i.e. fuse the results). Figure 6 illustrates this concept with an example from the acoustic processing of footsteps a noisy environment.

The example uses two complementary generic SPAs: a time-domain energy tracker and an STFT. The energy tracker detects a short uniform energy burst that should correspond to short tracks in the frequency domain. When analyzed by STFT-1 with its wide analysis window, the footstep's impulsive energy is smoothed with surrounding noise and fails to appear as frequency tracks in the STFT's correlates. The temporal locations and durations of the energy tracker's bursts serve two purposes. First, they indicate that STFT-1 was potentially inappropriate to the environment. Second, they serve as the basis for deciding where in the data stream to selectively apply STFT-2 with a narrower analysis window and smaller time decimation interval to find evidence for the potential new sound. The correlates of STFT-2 not only confirm the belief that STFT-1 was inappropriate to the environment, but also more strongly confirm the existence of the impulsive footsteps than the energy tracker's correlates did by themselves.

4. FUSION MECHANISMS

In this section we discuss the mechanisms we are developing for IPUS to effectively and efficiently fuse correlates from applications of various SPA instances. They consist of *discrepancy-detection tests* and data structures for representing *processing contexts*, and *context-mapping rules.*

Discrepancy-detection tests are a set of comparisons $\mathcal{C}$ defined on $\mathcal{F} \times \mathcal{F}$, where $\mathcal{F}$ is the set of generic front-end SPAs available to the IPUS system. Each test represents a heuristic consistency check that can be made between correlates of one generic SPA's instances and correlates from another generic SPA's instances. These tests are based on signal processing theory and are used to indicate when one SPA instance might no longer be capturing *all* desired features from the environment . For example, one test from the IPUS acoustic testbed compares the energy fluctuations from a time-domain energy tracker with the appearance of new peak tracks in the output of an STFT. It is theoretically valid in the sense that it is based on the requirement by formal processing theory

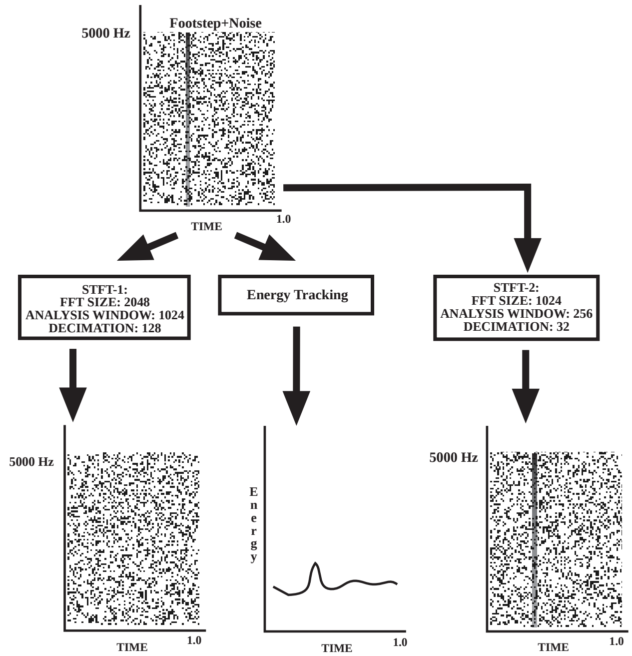

Figure 6: *Correlate comparison across different generic SPAs. The energy tracking SPA provides correlates for energy burst features. These features guide the focused application of an STFT with parameters to find frequency-track correlates for the footstep impulse in a noisy environment.*

that time-domain signal energy must be conserved in the frequency-domain. It is heuristic in the sense that not all time-domain energy "bursts" are due to the appearance of new sounds; sometimes they arise from interactions among the currently-active sources.

A processing context is a structure that stores relevant assumptions made by the IPUS system at the time a correlate was produced. In IPUS, every SPA correlate is tagged with a processing context. Specifically, the processing context contains:

1. the parameter context. This is the generic SPA whose instance produced the correlate and the values the SPA instance's parameters had at the time the correlate was produced.

2. interpretation assumptions about the signal. For example, what distortions like poor frequency resolution have already been identified.

3. the problem-solving goals in effect when the correlate were produced. For example, the goal of reducing uncertainty resulting from alternative interpretations for the same data, or the goal of finding correlates for a particular frequency track of a particular source.

4. the time period(s) for which the context is true.

By themselves, processing contexts provide a history of how the environment signal was processed, and with what results. Thus, they can support efficient re-use of previous reprocessings' results. By checking the parameter contexts of previously-created processing contexts, the IPUS reprocessing component can prevent re-execution of previously-selected reprocessing plans. It can also circumvent actual plan execution by checking for and re-using the correlates of earlier processing contexts with tighter parameter values than the current proposed plan. They are most useful for model synthesis, however, in connection with context-mapping rules.

Context-mapping rules are transformations defined on processing contexts. Each transform maps correlates computed by one instance of a generic SPA to their expected appearance if computed by a different instance of the SPA. The input to the transforms consists of observed correlates and their processing contexts, and the output is what correlates should appear under the new parameter context *assuming the other information in the original processing context is valid.* These transforms are useful in the IPUS reprocessing component in situations where correlates from two processings of the same signal must be compared to find evidence for a new object that is discernible only in the second processing's correlates. By mapping the context of the first processing into that of the second processing, we can eliminate from consideration those correlates in the second processing that correspond to previously-identified correlates in the first processing.

In our IPUS acoustic interpretation testbed we have one context-mapping rule for each generic SPA. Each rule manipulates one or more correlate features that can change under different processing contexts. The transform for the STFT covers feature changes including the merging and splitting of peaks as frequency and/or temporal resolution changes and the magnitude fluctuations that peaks can undergo due to beat phenomena.

To see the relationships among these mechanisms in IPUS, consider a scenario with a two-frequency source that can be modeled by

$$s(t) \quad = \quad \cos(2\pi 1200t) + \cos(2\pi 1220t) + f(t).$$

The scenario is sampled at 10 KHz, with f representing the rest of the acoustic environment, none of whose other components are closer to each other than 20 Hz. Assume futher that the scenario is being monitored by an STFT with a 1024-point analysis window. This scenario will have a "beat" of 10 Hz, or a period of 1000 data points over which the source's amplitude envelope will oscillate from 0.0 to at least 2.0 (see Figure 7). Assume at some time t an impulsive (approximately 0.1 sec duration) source appears. It will not be detectable in the STFT's correlates, but a discrepancy-detection test comparing time-domain energy tracking correlates and the STFT's correlates indicates its possible presence. Diagnosis indicates that spectral evidence for the impulse should be found by reprocessing data near time t with an STFT with a short (say 256 points) analysis window.

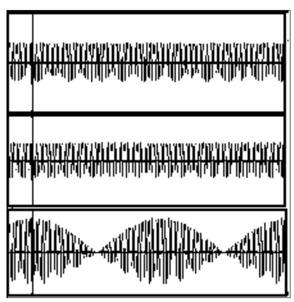

Figure 7: *A simple illustration of the beat phenomenon. The top two graphs show cosine waves at 1200 and 1220 Hz, respectively. The third graph shows the sum of the two cosine waves. All three graphs span 1100 data points, or 0.11 seconds. Note the induced 10 Hz beat in the third waveform.*

When the signal data was originally processed by an STFT with a 1024-point analysis window, an entire beat period was analyzed at a time, and the two sinusoids' magnitudes in the STFT instance's output spectra were relatively steady. When the data is reprocessed with a 128-point analysis window, however, the window's data will only cover an eighth of the beat period. Sometimes the window will include only the source signal's maxima and sometimes it will include only the source signal's minima, giving rise to wide variations in the observed magnitudes of the frequencies in the new STFT output. The variations can be so wide that a contouring method for identifying energy and frequency trends may mislabel the energy swings of these reprocessed versions of the source's contours as attack or decay behavior. Thus, while reprocessing is being performed to solve one discrepancy (the one caused by the impulsive source), it is inducing radically different behavior in previously-identified sources (e.g. data that gave rise to steady contours in a region now gives rise to attack or decay contours in the same region).

However, given the parameter context under which these "mislabelled" contours were created, and the parameter context into which support will be imported (the original context with a 1024 point analysis window), the reprocessing component can apply its STFT context-mapping rule to map the energy and frequency values of the new peaks into ranges in the target context and, if these ranges fall in frequency-energy regions where support is desired, the new contours can be reclassified as acceptable alternative views of the source (providing even more evidence for the source), not as new discrepancies to be diagnosed.

5. CONCLUSION

In this paper we have outlined a knowledge-based architecture for perceptual systems that makes no assumption about the appropriateness of its front-end SPAs to the current environment. The lack of this assumption turns perception into a dual search for appropriate SPAs and interpretations for an environment. In the course of these searches the signal data may be reprocessed several times by different SPAs with different parameter settings, raising the need for model synthesis, or the fusion of the reprocessings' results. We presented three domain-independent mechanisms for implementing this fusion: discrepancy detection tests, processing contexts, and context-mapping rules.

In future work we will examine the question of limited context-mapping. This refers to the possibility of mapping only some of the features of correlates from one context to another in order to conserve time.

6. ACKNOWLEDGMENTS

This work was supported by the Rome Air Development Center of the Air Force Systems Command under contract F30602-91-C-0038, and by the Office of Naval Research under contract N00014-92-J-1450. The content does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

7. REFERENCES

[1] Bitar, N.; Dorken, E.; Paneras, D.; and Nawab, H. "Integration of STFT and Wigner Analysis in a Knowledge-Based Sound Understanding System," *Proc. 1992 IEEE Int'l Conf. on Acoustics, Speech and Signal Processing.*

[2] N. Carver and V. Lesser, "A Planner for the Control of Problem-Solving Systems," to appear in *IEEE Transactions on Systems, Man, and Cybernetics* Special Issue on Planning, Scheduling and Control, 1993.

[3] N. Carver and V. Lesser, "A New Framework for Sensor Interpretation: Planning to Resolve Sources of Uncertainty," *AAAI '91 Proceedings*, pp. 724–731.

[4] Dawant, B.; Jansen, B. "Coupling Numerical and Symbolic Methods for Signal Interpretation," *IEEE Trans. Systems, Man and Cybernetics.* Jan/Feb 1991.

[5] Dorken, E.; Nawab, H.; Lesser, V. "Extended Model Variety Analysis for Integrated Processing and Understanding of Signals," *Proc. 1992 IEEE Int'l Conf. on Acoustics, Speech and Signal Processing.*

[6] Erman, L.; Hayes-Roth, R.; Lesser, V.; Reddy, D. "The Hearsay II Speech Understanding System: Integrating Knowledge to Resolve Uncertainty," *Computing Surveys*, v. 12, June 1980.

[7] Lesser, V.; Nawab, H.; Gallastegi, I.; Klassner, F. "IPUS: An Architecture for Integrated Signal Processing and Signal Interpretation in Complex Environments," *Proc. 1993 Nat'l Conf. on Artificial Intelligence.*

[8] Lesser, V.; Nawab, H.; Klassner, F. "IPUS: An Architecture for the Integrated Processing and Understanding of Signals," TR 93-29, Computer Science Dept., University of Massachusetts, Amherst, MA.

[9] Nawab, H.; Lesser, V. "Integrated Processing and Understanding of Signals," ch 6, *Knowledge-Based Signal Processing*, A. Oppenheim and H. Nawab, eds, Prentice Hall, New Jersey, 1991.

[10] Nawab, H.; Quatieri, T. "Short-Time Fourier Transform," *Advanced Topics in Signal Processing*, Prentice Hall, New Jersey, 1988.

[11] Nii, P.; Feigenbaum, E.; Anton, J.; Rockmore, A.; "Signal-to-Symbol Transformation: HASP/SIAP Case Study," *AI Magazine*, vol 3, Spring 1982.

[12] Swain, M.; Stricker, M. eds. *Promising Directions in Active Vision,* NSF Active Vision Workshop, TR CS 91-27, Computer Science Dept, University of Chicago, 1991.