Data Reprocessing and Assumption Representation in Signal Understanding Systems[†]

Frank Klassner
Computer Science Department
University of Massachusetts
Amherst, MA 01003 USA
klassner@cs.umass.edu

CS Technical Report 92-52
August 1992

Abstract

A key issue in the development of next-generation intelligent systems is the ability to perceive and understand the complex environments in which they will operate. Complex environments are characterized by variable signal-to-noise ratios, unpredictable source behavior, and the simultaneous occurrence of target sources whose signal signatures can overlap, mask, or otherwise distort each other. This paper argues that traditional perceptual architectures have limited effectiveness in such environments and presents an alternative design that is a significant extension of the *Integrated Processing and Understanding of Signals* (IPUS) architecture[*]. The IPUS philosophy emphasizes structured bidirectional interaction between numeric signal processing and symbolic interpretation processes. The interaction occurs as a result of search for signal processing control parameter values that produce evidence satisfying the interpretation processes' goals. This search is constrained by formal signal processing theory and dynamically generated problem-solving assumptions.

Within the overall goal of extending, generalizing, and validating the IPUS architecture, this research program will explore the utility and scalability of formally designing

[†] This paper is based upon work supported by the Rome Air Development Center of the Air Force Systems Command under contract F30602-91-C-0038. The content does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

[*] Under development in the Computer Science Department in the University of Massachusetts at Amherst.

perceptual systems with three features to make their processing strategies more adaptive to complex environments' demands: *Selective Processing Streams* (the ability to selectively apply signal processing algorithms to portions of the environment's signal data), *Multiple Views Synthesis* (the ability to integrate results from several front-end processings each with its own processing parameter-settings), and *Explicit Processing Assumptions* (the ability to represent and revise environmental and problem-solving assumptions as first-class objects and to use such information in all system components).

This paper presents the current status of the IPUS testbed, discusses issues to be addressed in extending its architecture to accommodate the new features, and describes how the new architecture should be evaluated.

1 Introduction and Motivation

A key issue recognized across the machine-perception subdisciplines [21, 34, 36] in the development of next-generation intelligent systems is the ability to perceive and understand the complex environments in which they will operate. Complex environments are characterized by variable signal-to-noise ratios, unpredictable source behavior, and the simultaneous occurrence of target sources whose signal signatures can overlap, mask, or otherwise distort each other. These observations place increased significance on intelligent systems' perceptual components and invite a critical examination of their current design paradigm.

In traditional perceptual systems [14, 30], the front-end signal processing is fixed, and interpretation processes are limited to analyzing only the single view afforded by that processing. "Fixed" front-end processing refers to the situation where signal processing algorithms (SPAs) are employed with fixed control parameter values. An example would be a speech recognition system whose front-end processing used a fixed-order linear-prediction algorithm with an analysis window length fixed to 512 points. This traditional design paradigm is based on two assumptions:

1. A small set of front-end fixed-parameter SPAs can produce evidence of sufficient quality for deriving plausible interpretations under all environmental scenarios.

2. A perceptual system's processing goals remain static with respect to the environment being monitored.

In restricted application domains with steady, relatively high signal-to-noise ratios, these design assumptions cause few problems. When traditional perceptual systems are used to monitor complex real-world environments, however, these assumptions lead very quickly to degraded interpretation quality.

The first assumption implies that traditional perceptual systems tend to ignore shortcomings of their front-end SPA-and-parameter sets with respect to unexpected changes in the monitored environment. For example, consider a system whose front-end processing consists of a Fast Fourier Transform (FFT) with 512-point analysis window and imagine that it is monitoring a sound source A whose two distinguishing frequency components are separated by 40 Hz. Assume that the system is supplied with a source-description database that in addition to A also contains a source B with a single frequency component in the same region as source A. If the signal is being sampled at 10 KHz, basic Fourier analysis theory [31] indicates that these

components' separation lies at nearly twice the limit of the front-end processing's frequency resolution capability. If the source's components should "drift" in frequency toward each other, the identifying components will appear to the system's interpretation processes as one merged component which could represent source A, source B, or both simultaneously (see figure 1). Source behavior changes, new source occurrences, and other environmental events can cause SPA results to appear unlike those expected for the monitored source because the front-end SPAs are not appropriate for the new environmental scenario. An SPA will be termed *appropriate* for a set S of signals if it produces output that meets all recognition requirements (e.g. all components adequately detected, all components resolved, etc) for each source-combination in 2^S.

Systems (e.g., [9, 23, 30, 37]) that partially support[1] the ability to recognize that signal data might have been processed by inappropriate front-end SPAs may incorporate this recognition in the uncertainty associated with their output's interpretations, but they tend to have limited formal strategies to resolve this source of uncertainty. As Carver [7] points out, most interpretation systems (of which perceptual systems are a subset) have limited means for resolving interpretation uncertainty because they are unable to generate an explicit record of the reasons for the uncertainty. The first design assumption places emphasis on developing front-end SPA sets that "get the right data the first time," which tends to be possible only by limiting application domains to stable, constrained environments. There has therefore been little motivation to provide such systems with robust capabilities to determine whether the front-end SPAs are still appropriate to the environment (i.e., detect uncertainty), to explain why they are not appropriate if they have been found to be so (i.e., generate reasons for the uncertainty), and to use the explanation to modify interpretations, their certainties, or the front-end processing itself.

The second assumption leads to a conceptual "disconnection" between the application of front-end SPAs and the perceptual system's dynamic processing goals. It discourages the selective, goal-directed application of specialized (often low-cost) SPAs to provide only enough data to resolve specific uncertainties engendered by changes in the processing goals due to changes in the monitored signal. For example, a system's primary goal might be to respond to either the sounds of an infant or a ringing telephone and to ignore other sound sources. This may be done by moni-

[1] "Partially" refers to the use of catchall interpretation hypotheses such as UNKNOWN-SOURCE or catchall distortion explanations such as RANDOM-SOURCE-NOISE or the use of clever probabilistic weighting techniques on certainty factors without explicit identification of (1) source(s) the data could actually represent *and* (2) environmental factors that led to the distortion.

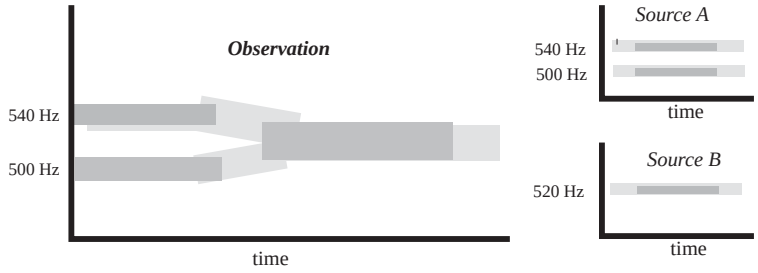

Figure 1: *An example of ambiguity introduced by fixed front-end processing. When source A's components "drifted" toward each other, the FFT frequency resolving capability is exceeded and they appear merged as one frequency track which could represent a poorly processed source A, the appearance of source B and the disappearance of source A, or the simultaneous presence of source A and source B. Heavy shading indicates high signal energy, lighter shadings indicate low signal energy.*

toring all frequencies from a series of suitably-configured FFTs. If an infant sound is detected, the system's goal may then switch to determining quickly whether the infant is crying or choking while ignoring telephone rings. Such a goal might be accomplished by temporarily monitoring only a few low-frequency spectral regions with a specialized, low-cost SPA such as Goertzel's algorithm [16][2]. The traditional perceptual system design could implement such SPA switching within its fixed-front-end framework by applying each available SPA (both specialized and general) to the signal data and selectively examining the SPA-output streams when necessary. The traditional paradigm's view of signal processing as non-adaptive is inappropriate for applications where the complex, situation-dependent nature of signal processing requirements leads to a combinatorial explosion in the number of different SPAs that a signal understanding system must have at its disposal.

Thus, traditional perceptual system architectures are ill-equipped for monitoring complex environments because their front-end processing is unresponsive to environmental changes and provides little help beyond its fixed view in resolving ambiguous data. This paper provides an initial exploration of the issues to be addressed in formally designing perceptual systems with three features to make their processing strategies more adaptive to the demands of complex environments:

[2]It can be shown ([31], ch 9) that Goertzel's algorithm is faster than the FFT when fewer than $\log N$ frequency samples are required, with N being the number of signal sample points in one analysis window.

Selective Processing Streams: the ability to selectively apply SPAs to portions of the environment's signal data.

Multiple Views Synthesis: the ability to integrate results from several processings under a variety of SPA parameter-settings.

Explicit Processing Assumptions: the ability to represent and revise environmental and problem-solving assumptions as first-class objects and to use such information in all system components.

The first two features are intended to overcome the generality shortcomings of the traditional paradigm assumptions as well as to provide tools for resolving interpretation uncertainties. The third is intended to improve traditional perceptual systems' ability to represent and respond to the reasons for their interpretations' uncertainty. Although this paper will focus on perception in the auditory modality, it should be noted that research toward adaptive front-end signal processing has also been recognized in the active computer vision community as an important next step in machine perception [36].

The *selective processing streams* capability is important to signal interpretation systems because it would permit the resolution or significant reduction of uncertainties associated with the application of SPAs inappropriate for monitoring the current state of the sources in a dynamic environment. When the current front-end processing configuration produces ambiguous data with too many alternative interpretations, selective processing of previous portions of the signal with different SPA parameter settings or with specialized SPAs can significantly prune the interpretation search space to be considered. We will refer to the re-examination of ambiguous data with different SPAs alternatively as *reprocessing* or *local parameter adaptation.* The ability to reprocess signal data eases the front-end processing design burden of choosing many SPAs (some quite time-consuming) to "get the data right" on the first encounter. It permits system designers to use cheaper, less accurate SPAs "most of the time," knowing that short periods of ambiguous SPA output can be handled by limited reprocessing of the signal by more precise, costlier SPAs. Selective processing streams also would support *global parameter adaptation*, which refers to changing the default front-end SPAs and/or their parameter settings in anticipation of future signal data characteristics. This anticipation can arise from several reasons: monitored sources have unexpectedly changed behavior and cannot be monitored by current front-end processing with adequate certainty; source models indicate upcoming source behavior changes that will not be observable by

current front-end processing; too much reprocessing of buffered data is being done which should have been done *once* as front-end processing; etc.

In [27] the *multiple views synthesis* capability is described as a desirable feature for next-generation signal interpretation systems because it could permit the fashioning of interpretations from evidence obtained from disparate processings. It is often the case that sufficient support evidence for an interpretation hypothesis can only be gathered by analyzing the raw signal data under several sets of SPA parameter values, each of which precludes the availability of the others' evidence. Consider, for example, the examination of signal data for a sound source with a pair of synchronized frequency components which have short signal onset times (say, 0.2 sec) and are separated in frequency by 15 Hz. Assuming that the signal was sampled at 10KHz, a series of 2048-point FFT algorithms applied to consecutive 2048-point analysis windows in the data will provide adequate frequency resolution to detect both microstream. The short onset period's energy variation will not be detectable, however, because the analysis windows are too wide and "blend" the energy of the 1000 points associated with the onset with that of the surrounding points. A series of FFTs with shorter analysis windows (say 512 points) could provide us with sufficient evidence to conclude the attack region was of the specified length, but it would not detect both microstreams. The shorter windows would limit the FFTs' frequency resolving power and merge the two microstreams into one. Thus, two properties (onset duration and frequency components) of a single source hypothesis required two different views of the signal data for confirmation. See figure 2 for illustration.

The presence of *environmental and problem-solving assumptions* as first-class objects available to all system components is an important addition to signal interpretation systems. It provides the basis for more sophisticated diagnostic strategies in determining why a signal expectation was violated (e.g., reasons for interpretation uncertainty [7]) and whether or not to reprocess data to find new corroborating data for the explanation. If, for example, an interpretation system has access to the assumption that there are no new sound sources that have appeared in the environment, then a diagnostic subprocess in the system could eliminate the possibility of a newly-occurring source masking a previously-observed source as an explanation for any "fadeout" or loss of signal detected in the observed source. Access to explicit processing goals would also provide a formal basis for signal interpretation systems to select new front-end processing strategies when the environment necessitates changes in processing goals. A history of the problem-solving goals that motivated data reprocessing under new parameters can serve as an indicator for

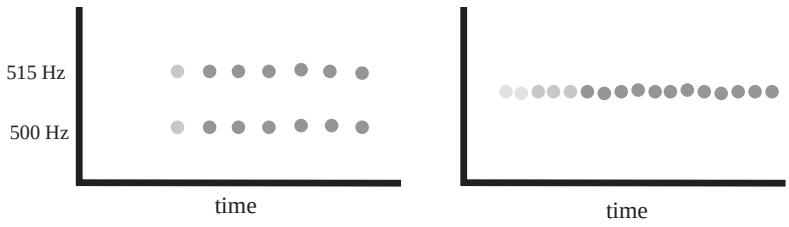

Figure 2: *Spots indicate detected frequency peaks. The left figure shows the result of successive 2048-point FFT applications, with peaks separated by about 0.2 seconds. The right figure shows the result of successive 512-point FFT applications on the same signal, with peaks separated by about 0.05 seconds. Heavy shading indicates high signal energy, lighter shadings indicate low signal energy.*

when too much reprocessing is occurring that could be reduced by global parameter adaptation. In facilitating control of reprocessing, explicit assumptions such as source behavior models can also serve to suppress reprocessing by predicting distorting interactions (e.g. source masking) and identifying distortions as acceptable evidence given that they match the predictions.

Research on the three capabilities will be carried out within the IPUS (*Integrated Processing and Understanding of Signals*) sound understanding testbed developed in the Computer Science department at the University of Massachusetts at Amherst [22]. The testbed's fundamental philosophy is that the search for SPAs appropriate to an environment's dynamic state must interact with the search for correct SPA output interpretations in a structured bi-directional manner.

The following is a sketch of the IPUS framework's primary components and their relationships. A *discrepancy detection* mechanism checks for discrepancies between front-end SPA output and (1) the output's expected form, (2) alternative SPAs' output, and (3) application-domain signal characteristics. If discrepancies are detected, a *diagnosis* process is then executed to obtain a mapping from the discrepancies and SPA parameter values under which they were observed to qualitative hypotheses that explain the distortions. This process uses the *formal theory* underlying the signal processing. A *signal reprocessing* process then proposes and executes a search plan to find a new set of values for the generic SPA(s)' parameters that eliminates or reduces the hypothesized distortions. During this plan's execution, the signal data may be reprocessed several times under different SPAs with different param-

eter values. Each time the data is reprocessed, a new parameter-value state in the SPA parameter search space is examined both for how well it eliminates or reduces distortions and for how well the new output supports the original interpretation. Reprocessing is also used by a *differential diagnosis* process to reduce uncertainty when data supports several plausible interpretations with relatively equal strength. Diagnosis differs from differential diagnosis in that the former attempts to explain why a particular desired situation is not occurring, while the latter attempts to differentiate several equally plausible situations.

This paper's research will make several contributions to the field of perceptual system design:

1. A generalized architecture formally integrating the key IPUS processes of discrepancy detection, diagnosis, reprocessing, and differential diagnosis with explicit processing assumptions, selective processing streams, and multiple views synthesis. As will be seen in section 3, the current architecture is missing a global parameter adaptation mechanism, its existing key processes are not rigorously formalized, and the interface among the processes is still incompletely specified.

2. An answer to the problem of how the multiple views afforded by selective processing streams can be exploited by problem-solving strategies augmented with explicit processing assumptions.

3. A demonstration of the architecture's applicability and generality.

4. A quantitative measure of the reprocessing capability's importance to signal interpretation quality.

5. A viable platform for future experimentation on psychoacoustic streaming theories [5] and control strategies for acoustic signal interpretation.

This paper (1) presents related work, (2) describes the IPUS paradigm and the current state of the testbed in which the research will be carried out, (3) describes mechanisms for implementing the three new capabilities in the testbed and the issues to be faced while they are integrated with existing testbed components, and (4) discusses an evaluation framework for the research.

2 Related Work

Several recent systems have been developed in the exploration of frameworks for structured interaction between interpretation activity and numeric-level signal processing. Each has made use of components performing at least one of the tasks of the IPUS testbed components, but none have included (and therefore none have attempted to interface) components for *all* the IPUS components' tasks. The following descriptions will illustrate recent frameworks, their differences from the IPUS framework, and the interpretation limitations these differences impose on them in order to justify a call for an examination and formalization of the relationships between the current and proposed IPUS components: discrepancy detection, discrepancy diagnosis, reprocessing, and differential diagnosis, explicit processing assumptions, reprocessing and multiple views synthesis.

Hayes-Roth's GUARDIAN system [17] incorporates an input-data management component that controls the sampling rate of signals in response to overall system workload constraints. The framework is typical of systems whose input data points already represent useful information and require little formal front-end processing other than to control the rate of information flow. Information flow is controlled through variable sample-value thresholds and variable sampling rates. This interaction framework is somewhat limited since it is based only on system reasoning-time requirements, and provides good performance primarily because the signals monitored are relatively simple and noise-free in nature: heart-rate, temperature fluctuations, etc. The framework does not appear adequate for the general class of signals IPUS is to encounter: signals containing complex structures that must be modeled over time in the presence of variable noise levels.

Dawant's framework [9] is more general and separates signal interpretation knowledge from signal processing knowledge. It also supports the concept of multiple SPA views (called *channels* in his terminology). However, the framework does not support the selective processing stream concept since data is always gathered on every channel whether required for interpretation improvement or not. System control appears highly goal-directed and employs a limited representation of model uncertainty (only three levels of certainty to characterize data matches with signal event models). Framework descriptions make it appear that, unlike IPUS, it operates on the implicit assumption that the signal-generating environment will not interact adversely with the signal processing algorithms' limitations to produce output distortions that might not have occurred if more appropriate processing algorithms had been used. Any deviations between observed signal behavior and available sig-

nal event models are attributed to chance variations in the *source* being monitored, never to the signal's *interaction* with inappropriate SPAs or with other sources in the environment.

In GOLDIE [20], Kohl describes an image segmentation system that permits high-level interpretation goals to guide the choice of numeric-level segmentation algorithms, their sensitivity settings, and region of application within an image. The system can engage in a "hypothesize-and-test" search strategy for algorithms that will satisfy high-level goals, given the current image data. While it incorporates an explicit representation of algorithm capabilities to aid in this search, and an explicit representation of reasons for why it assumes an algorithm is appropriate or inappropriate to a particular region, the system does not incorporate a centralized diagnosis component for analyzing unexpected "low quality" segmentations. If an algorithm were applied to a region and the resulting segmentation were of unexpectedly low quality, the framework would not parallel IPUS and attempt to diagnose the discrepancy and exploit this information to reformulate the algorithm search but would select the next highest rated algorithm and proceed.

De Mori et al. [11] developed a formal interaction framework in a system to recognize letters of the English alphabet. Interpretations were generated by learned rules expressing letter identifications in terms of a signal-event grammar. As an example, the letter 'V' may be present when a short deep dip in the signal's time-domain energy is followed by a long peak in time-domain energy. Often more than one letter can be indicated by a single rule (in their terminology the rule has a *confusion set*). When such rules are activated, the system pursues a differential diagnosis strategy relying on rules describing SPAs that are suited to disambiguating confusion sets with given members. Thus, the system makes use of the selective processing stream concept and differential diagnosis strategies. However, given the framework's relatively restricted application domain, there is a serious question of whether the approach can be scaled up without including the ability to model the environment in explicit assumptions. Since the system considers letters as isolated, unrelated words, the framework does not incorporate any use of diagnosis in conjunction with environmental constraints (e.g. A 'C' has been identified at time t_{-1} and a 'B' is expected at time t_0 since there is an environmental constraint that 'B's follow 'C's. No behavior supporting the expectation is observed, so diagnostic reasoning should be attempted to explain why).

Bell and Pau [2, 3] have formalized the search for processing parameter values in numeric-level image understanding algorithms in terms of the Prolog language's unification and backtracking mechanisms. They express an SPA as a predicate

defined on tuples of the form $(M, p_1, \ldots, p_n)$, where M represents an image pattern and the p's represent SPA control parameters. The predicate is true for all tuples where M can be found in the SPA output when its control values are set to the tuple's p values. Prolog's unification mechanism enables these predicates to be used in both goal-directed and data-driven modes. In a goal-driven mode, M is specified and some[3] of the parameters are left unbound. The unification mechanism seeks to verify the predicate by iteratively binding the unspecified parameters to values from a permissible value set, applying the SPA, then checking if the pattern is found. In a data-driven mode, M is not bound and the parameter values are set to those of the front-end processing. After the SPA is applied, M is bound to the results.

The method relies on Prolog's backtracking *cuts* [15] to limit parameter-value search. A cut is a point in the verification search space beyond which Prolog cannot backtrack. This reliance on a language primitive makes it difficult to explicitly represent (and therefore to reason about) heuristic expert knowledge for constraining parameter-value search as can be done in IPUS's reprocessing component. The cut mechanism also does not permit the use of diagnostic reasoning to further constrain parameter-value search based on the cause of an SPA predicate failure.

Multiple views synthesis is related to the problem of sensor fusion [25] in that both tasks are concerned with the combination of evidence from different views of the same entity to produce a single data structure representing an interpretation system's perception of the entity. However, sensor fusion has traditionally considered "view" to refer to data supplied by a sensor with fixed control parameters and "combination" to refer to the process of pooling evidence (data) from sensors in different modalities. Multiple views synthesis considers "view" to refer to data obtained under one set of processing control parameters and "combination" to refer to the process of integrating data in the same modality produced under different sets of control parameters. This distinction is not an attempt to present multiple views synthesis as some radical concept; rather, it is intended to illustrate this research's specific focus within the sensor fusion problem domain.

Given the adaptive nature of the IPUS architecture that will be presented in section 3, it is important to distinguish between the IPUS approach and the classic adaptive control theory approach [35]. Control theory uses stochastic-process concepts to characterize signals, and these characterizations are limited to probabilistic moments, usually no higher than second-order. Discrepancies between these

[3]Or possibly none, in which case no unification-directed search actually takes place; the pattern's presence is checked only for the one set of parameter-values.

stochastic characterizations and an SPA's output data are used to adapt future signal processing. In contrast, the IPUS architecture uses high-level symbolic descriptions (i.e. interpretation models of individual sources) as well as numeric relationships between the outputs of several different SPAs to characterize signal data. Discrepancies between these characterizations and SPAs' output data are used to adjust future signal processing. Classic adaptive control should therefore be viewed as a special case of an IPUS architecture, where the interpretation models are described solely in terms of probabilistic measures and low-level descriptions of signal parameters.

IPUS uses Carver's RESUN [6] framework to control knowledge source (KS) execution. This framework views interpretation as a process of gathering evidence to resolve particular hypotheses' sources of uncertainty (SOU) in the interpretation hypotheses. It incorporates a symbolic language for representing SOUs. The SOUs are structures used by system control mechanisms to select appropriate interpretation strategies. Problem-solving is driven by the information maintained in a *problem solving model*, which provides a summary of the current interpretation of data as well as a summary of the SOUs associated with each high-level hypothesis. An incremental, reactive planner maintains control using *control plans* and *focusing heuristics*. Control plans are schemas that define the interpretation methods and information gathering actions (e.g., SPAs) available to the system for processing and interpreting data, and for resolving interpretation uncertainties. Focusing heuristics select SOUs to resolve and processing strategies to pursue when there are several possibilities. In this way high-level interpretation SOUs such as missing support, alternative explanations, etc., can trigger reprocessing to reduce them to acceptable levels.

The RESUN framework was developed to address current interpretation systems' limited ability to express and react to the reasons for interpretation hypotheses' uncertainty. It emphasizes the separation of hypothesis belief evaluation from control decision evaluation by making control responsive to the presence of SOUs in the problem-solving model, not to the levels of belief in existing hypotheses. The control plan formalism is general enough to support both differential diagnosis and discrepancy diagnosis reasoning, and permits reprocessing strategies to be expressed as alternative control plans to be selected on the basis of SOUs describing discrepancies and their explanations. However, the framework does not preserve a history of the problem-solving model, nor does it provide mechanisms to record the control plans and goals that precipitated the gathering of a particular piece of evidence. Such historical records are important to providing a processing context on which to base the more sophisticated diagnostic reasoning envisioned for IPUS. It also

does not support adequate data maintenance to keep the system from exhausting available memory and "choking" on a full database.

3 The IPUS Paradigm

3.1 Architecture Summary

The starting point of the IPUS architecture design is its SPA database. The database contains a generic SPA for each algorithm class available to the IPUS system. An SPA instance is specified by values for a generic SPA's parameters, and has capabilities and limitations stemming from those values.

As an illustration, consider the Short Time Fourier Transform (STFT) algorithm class [28]. An instance in this class results from particular values for its parameters, such as window length (number of data points analyzed at a time), frequency-sampling rate, temporal decimation factor (consecutive analysis window overlap), etc. The instances differ from each other because their parameter values imply different assumptions about the input signal's spectral features and their time-variant nature. Instances with large window lengths may provide fine frequency resolution for signals whose frequencies remain steady over time, but at the cost of poor time resolution for signals whose components quickly shift within the frequency spectrum over time.

The IPUS architecture's basis is an iterative search technique for converging to the appropriate SPAs and parameter values. The following summarizes the technique; later sections and [22] provide a more detailed view of its components.

The technique starts with a best guess for front-end SPAs and parameter values to process the input signal (an arbitrary set is used in the absence of environmental knowledge). A *discrepancy detection* mechanism then checks for significant discrepancies between front-end SPA output and

1. the output's expected form,
2. alternative SPAs' output, and
3. application-domain signal characteristics.

Significant discrepancies are those which, if left unresolved, will lead to incorrect interpretations or to large amounts of time spent in extra interpretation search. If such discrepancies are detected, *diagnosis* is then performed to obtain a mapping

from the discrepancies and SPA parameter values under which they were observed to qualitative hypotheses that explain the distortions. This process uses the formal theory underlying the signal processing. A *signal reprocessing* stage then proposes and executes a search plan to find a new set of values for the generic SPA(s)' parameters that eliminates or reduces the hypothesized distortions[4]. During the plan's execution, the signal data may be reprocessed several times under different SPAs with different parameter values. Each time the data is reprocessed, a new parameter-value state in the SPA parameter search space is examined for how well it eliminates or reduces distortions. Reprocessing is also used to perform *differential diagnosis* to reduce uncertainty when data supports more than one plausible interpretation. Differential diagnosis is a process whereby certain characteristics of the subject being analyzed (in this case, a signal) are methodically highlighted or exaggerated to make it possible to more clearly distinguish which one of several interpretations of the subject is the most plausible.

It might appear that the IPUS paradigm's reliance on data reprocessings to reduce interpretation uncertainty could be criticized for the reprocessing time cost. Such criticism is not appropriate, since as hardware advances, SPA time costs keep decreasing. Additionally, in a traditional system each algorithm (specialized, and/or just many instances of STFT) would always be applied to the data stream, producing K processing output streams. In addition to the time spent generating all of this data, one must also consider the time consumed in analyzing and integrating this data with existing interpretation hypotheses. *This* truly represents an unacceptable time cost. What IPUS permits is the selective creation and sampling of each processing output stream, actually cutting down the resources required by K continuously-sampled output streams. So rather than view IPUS as an approach that adds time costs to traditional approaches, one should view IPUS as an approach that cuts time costs from indiscriminate application of traditional approaches to complex scenarios.

IPUS is designed to serve as the basis of perceptual systems that are driven by the goal of producing interpretations with acceptable uncertainty levels. Therefore, control in IPUS requires a formalism for representing factors that affect their interpretations' confidence levels. The control mechanism must also be able to focus on particular uncertainties in a context-sensitive manner.

For these reasons, IPUS uses the RESUN [6] framework to control knowledge

[4]This three-step process is similar to the one developed in [19] for meta-level control in problem solving systems.

source (KS) execution. This framework views interpretation as a process of gathering evidence to resolve particular hypotheses' sources of uncertainty (SOUs) in the interpretation hypotheses. It incorporates a symbolic language for representing SOUs. The SOUs are structures used by system control mechanisms to select appropriate interpretation strategies. Problem-solving is driven by the information maintained in a *problem solving model*, which provides a summary of the current interpretation of data as well as a summary of the SOUs associated with each high-level hypothesis. An incremental, reactive planner maintains control using *control plans* and *focusing heuristics*. Control plans are schemas that define the interpretation methods and information gathering actions (e.g., SPAs) available to the system for processing and interpreting data, and for resolving interpretation uncertainties. Focusing heuristics select SOUs to resolve and processing strategies to pursue when there are several possibilities. In this way high-level interpretation SOUs such as missing support, alternative explanations, etc., can trigger reprocessing to reduce them to acceptable levels.

Figure 3a shows the generic IPUS architecture, while figure 3b shows the architecture's instantiation in the sound understanding testbed.

3.2 Current IPUS Testbed Status

This section summarizes the state of the IPUS sound understanding testbed as background for section 4's discussion of the research and testbed enhancements advocated by this paper. It represents the work done by the IPUS research group[5] since September 1989. This summary covers the testbed's knowledge representations, its major knowledge sources, and their relationships. Refering to figure 3, the primary IPUS tasks are

1. Discrepancy Detection
2. Discrepancy Diagnosis
3. Reprocessing
4. Differential Diagnosis.

[5] Victor Lesser, Hamid Nawab, Malini Bhandaru, Zarko Cvetanović, Erkan Dorken, Izaskun Gallastegi, and Frank Klassner.

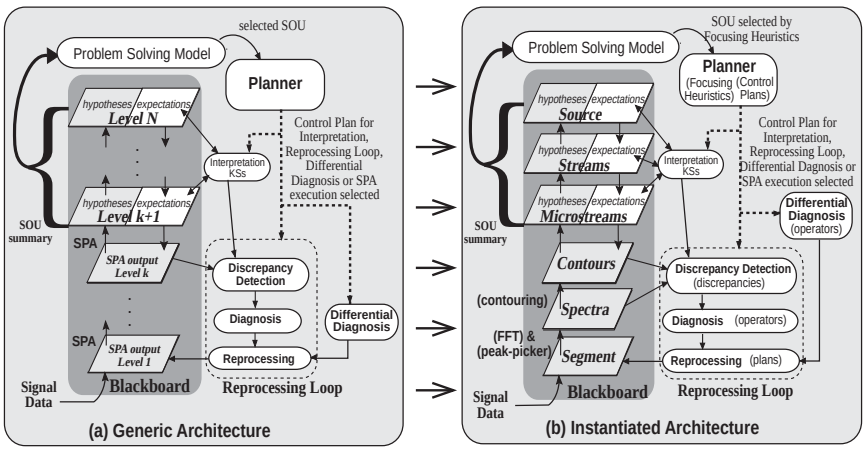

Figure 3: *3a shows the generic IPUS architecture, 3b shows the architecture instantiated for the sound understanding testbed. Solid arrows indicate dataflow relations. Dotted arrows indicate plans that the planner can pursue when trying to reduce SOUs (discrepancies) in the problem solving model that were selected by the focusing heuristics. Knowledge to instantiate the architecture for an application is shown in parentheses in 3b. Reprocessing plans can produce SPA output at any abstraction level, not just the lowest.*

The testbed consists of a blackboard with seven evidence abstraction levels, KSs for the primary tasks and for inferring hypotheses between different abstraction levels, an acoustic source library, and control plans. Figure 4 shows the support relationships among the evidence abstraction levels, while figures 5 to 11 provide a short description of the information represented in the evidence abstractions.

At the lowest level are waveform segments derived from the input waveform. Each segment is a collection of points to which some SPA will be applied. The second level consists of spectral hypotheses derived for each waveform segment through frequency- and time-domain SPAs. The third level consists of "contour" hypotheses, each of which corresponds to a group of peaks (each from a different segment) whose time indices, frequencies, and amplitudes represent a contour in the time-frequency-energy space with uniform frequency and energy behavior. The fourth level contains microstream hypotheses supported by one contour or a sequence of contours. Each microstream has an energy pattern consisting of an attack region (signal onset), a steady region, and a decay (signal fadeout) region. In the fifth

level we represent noisebeds as wideband frequency regions supported by clusters of relatively-low energy contours. Noisebeds represent the wideband component of a sound source's acoustic signature. Clusters represent groups of contours for which no microstream or noisebed hypothesis can be generated given the inference KS's default interpretation criteria (e.g., noisebed contours must have an energy correlation of greater than X, microstream contours should not be separated from each other by more than Y milliseconds, etc). Groups of microstreams and noisebeds synchronized according to time and/or some other psychoacoustic criteria such as harmonic frequency sets support "stream" hypotheses in the sixth level. For a detailed description of acoustic streaming processes, see [5]. At the seventh level, sequences of stream hypotheses are interpreted as sound-source hypotheses.

Sources are represented in the source database by an acoustic grammar specifying microstream and noisebed frequency ranges and permissible ranges of energy relationships among microstreams and noisebeds within source streams. The grammar also specifies the permissible range of durations for each source's microstreams and streams, and the stream sequences and periodic patterns that characterize the source. For greater flexibility, each source has its own set of evidence combination tables defining the relative importance of each source substructure (e.g., microstream regions and streams) to the belief assigned to the superstructure it supports.

3.2.1 Discrepancy Detection KS

The discrepancy detection KS is crucial to the IPUS framework's iterative approach. It not only detects discrepancies but also categorizes them to permit a choice of actions based upon their severity or importance to the current processing context. The idea behind IPUS's current categorization is that when SPA output data is distorted, a signal understanding system must be able to detect discrepancies between this data and one or more of the following:

1. the *expected* SPA output based on source models. There are two classes of model-based expectations. The first is the set of models for sources already assumed to be present. The second is the set of models for sources under consideration for interpreting newly-detected evidence in the current block of data. Conflict discrepancies may involve either a total or a partial mismatch between evidence and the hypotheses it was supposed to support. An example of a total conflict occurs when the interpretations of past data show two sinusoids at 200 Hz and at 250 Hz with no decline in their amplitudes and

the current SPA output data contains neither of the sinusoids. A case where a partial conflict would be raised is when current data contained two out of three frequencies that supported the identification of a telephone ring, and after a search for the other frequency the system couldn't find it.

2. the *output data* from other signal processing algorithms applied to the same signal data. Such discrepancies are termed *faults*. For example, suppose that the signal data is being processed with a zero-crossing analyzer and an STFT. If the zero-crossing analyzer were to indicate the presence of a sinusoidal signal but the STFT does not, a fault would be declared. For another example see Bitar et al. [4], which describes an algorithm for comparing Wigner-Distribution and STFT spectra to detect faults.

3. the application domain's *a priori signal constraints*. Such discrepancies are termed *violations*. A violation occurs when the SPA output data has characteristics that are known to be absent in the entire class of possible signals in the application domain. For example, if the application domain is known to consist only of signals with frequencies below 500 Hz, SPA output data showing a signal at 700 Hz would give rise to a violation.

Conflict discrepancy detection is distributed among all the KSs responsible for interpreting lower-level data as higher-level concepts. Each such KS, when acting in a goal-directed manner, checks if any data can support the sought-after expectation. If none can be found, or if only partially supportive data is available, the KS will record this as an SOU in the problem solving model, to be resolved at the discretion of the focusing heuristics. At the end of each data block's numeric signal processing, an SPA discrepancy detection KS checks if SPA outputs are consistent with each other, testing for violations and faults.

An important consideration in discrepancy detection is that expectation hypotheses are often qualitative, as the example "within the next two seconds, a sinusoidal component currently at 1200 Hz will shift to a frequency between 1200 and 2000 Hz and the shift will last between 100 and 500 msec" shows. This implies that discrepancy detection mechanisms must be able to work with ranges as well as specific values. This requires a representation in which qualitative calculus can be performed. The range calculus used in the sound understanding testbed is similar to Allen's [1] common-sense temporal theory.

3.2.2 Discrepancy Diagnosis KS

The discrepancy diagnosis KS "explains" discrepancies between expected signal behavior and observed signal behavior. It models experts' use of SPA Fourier theory in diagnostic reasoning with a means-ends analysis framework incorporating multiple abstraction levels and a verification phase [29]. The KS accepts two inputs: an *initial state* representing expected signal behavior (expectation hypotheses and/or less precise but more reliable outputs from other SPAs) and a *goal state* representing the observed front-end SPA output. The initial state can include qualitative signal feature descriptions to handle uncertain and approximate information. The KS's formal task is to generate a "distortion operator" sequence mapping the initial state onto the goal state. The KS has a database of operators that model distortions resulting from improper SPA parameter values. For example, one operator models an STFT SPA frequency-resolution distortion related to the SPA's window-length parameter value, which occurs in the SPA output when the window-length parameter W and signal sampling rate R interact to cause frequency components closer than R/W to appear merged in the STFT output.

The KS's search for an explanatory distortion operator sequence is carried out using progressively more complex abstractions of the initial and goal states, until a level is reached where a sequence can be generated using no more signal information than is available at that level. Thus, the KS mimics expert diagnostic reasoning in that it offers simplest explanations first [33]. Once a sequence is found, the KS enters its *verify* phase, "drops" to the lowest abstraction level available for the initial state, and checks that each operator's pre- and post-conditions are met when all available state information is considered. If verification succeeds, the operator sequence and a diagnosis region indicating the signal structures involved in the discrepancy are returned. If it fails, the KS attempts to "patch" the sequence by finding operator subsequences that eliminate the unmet conditions and inserting them in the original sequence.

One issue not addressed in [29] that arises in the IPUS framework is the problem of inapplicable explanations. Because of the KS's preference for short explanations, sometimes the first explanation offered by the KS will not enable the reprocessing mechanism to eliminate a discrepancy. In these cases, the architecture permits reactivation of the diagnostic KS with the original explanation supplied as one that must not be returned again. To prevent repetition of the search performed when generating the original explanation, the diagnosis KS stores with the explanation the search-tree context it was in when the explanation was produced. The KS's

search for a new distortion operator sequence begins from that point.

3.2.3 Reprocessing KS

Once distortions have been explained, it falls to the reprocessing KS to search for appropriate SPAs and parameter values that can reduce or remove them. This component incorporates the following phases: *assessment, plan selection, and plan execution*. The input to the reprocessing KS includes a description of the input and output states, the distortion operator sequence hypothesized by the diagnosis KS, and a description of the discrepancies present between the input and output states. The assessment phase uses case-based reasoning to generate multiple reprocessing plans, each of which has the potential of eliminating the hypothesized distortions present in the current situation.

In the plan selection stage, a plan is selected from the applicable plan set based on computation costs or criteria supplied by control plans. The plan execution phase consists of incrementally adjusting the SPAs parameters, applying the SPAs to the portion of the signal data that is hypothesized to contain distortions, and testing for discrepancy removal. The process is necessarily incremental because the situation description is at least partially qualitative, and therefore it is generally impossible to predict *a priori* exact parameter values to be used in the reprocessing.

Reprocessing continues until distortion removal is achieved or plan failure is noted. Plan failure is indicated when either the number of plan iterations exceeds a fixed threshold or a plan iteration requires a SPA parameter to have a value outside fixed bounds. When failure occurs, the diagnosis KS can be re-invoked to find an alternative explanation for the original distortions. If no alternative explanation can be found, an IPUS system annotates the hypotheses involved in the discrepancy with SOUs indicating low confidence due to unresolvable discrepancies.

3.2.4 Differential Diagnosis KS

The differential diagnosis KS produces reprocessing plans that will enable the system to prune the interpretation search space when ambiguous data is encountered. Its input is the ambiguous data's set of alternative interpretations, and it returns the time period in the signal to be reprocessed, the support evidence each interpretation requires, and the set of proposed reprocessing plans.

The KS first labels any observed evidence in the interpretation hypotheses' overlapping regions as "ambiguous". It then determines the hypotheses' discriminating

regions. For each discriminating region with no observed evidence, the KS posits an explanation for how the evidence could have gone undetected, assuming the source was present. These explanations index into a plan database, and select reprocessing plans to cause the missing evidence to appear. The KS then checks each ambiguous data region for resolution problems based on source models (e.g., a frequency region's data could support one source Y component or two source Z components), and selects reprocessing plans to provide finer component resolution in those regions.

The reprocessing plan set returned is the first non-empty set in the sequence: missing-evidence and ambiguous-evidence plan sets' intersection, missing-evidence plan set, ambiguous-evidence plan set. This hierarchy returns the plans most likely to prune a large number of interpretations from further consideration. The region of mutual temporal overlap for the alternative hypotheses defines the reprocessing time region, and the ambiguous and missing evidence handled by the reprocessing plan set defines the support evidence. A plan from the returned set is then iteratively executed as in the reprocessing KS until either a plan-failure criterion is met or at least one support evidence element is found.

This KS's explanatory reasoning for missing evidence is primitive compared to the discrepancy diagnosis KS's. Only simple, single distortions such as loss of low-energy components due to energy thresholding are considered; no multiple-distortion explanations can be constructed. This design is justified given that the KS's role is to quickly prune large areas of interpretation spaces, *without preference* for any particular interpretation. When a particular interpretation is rated over alternatives and an explanation for its missing support is required, an IPUS system uses the discrepancy diagnosis KS, encoding the preferred interpretation in the initial state.

4 Proposed IPUS Additions and Issues

This section describes supporting structures and mechanisms proposed for formally realizing the following features and outlines the issues that must be addressed in implementing them and interfacing them with existing IPUS testbed components.

Selective Processing Streams: the ability to selectively apply SPAs to portions of the environment's signal data.

Multiple Views Synthesis: the ability to integrate results from several processings under a variety of SPA parameter-settings.

Explicit Processing Assumptions: the ability to represent and revise environmental and problem-solving assumptions as first-class objects and to use such information in all system components.

This introductory subsection emphasizes and justifies five new components: processing contexts, a context-mapping KS, context-switching cost graphs, a source-model reconfiguration KS, and a global-parameter-adaptation KS. The subsequent subsections discuss implementation and intellectual issues that arise with respect to each feature as the new features are integrated with the testbed. The reader should keep in mind that although time-related issues will occasionally appear in the following discussion, they should not be taken to indicate that this research plan will attempt to formally address real-time issues per se. Whenever time issues are considered, they will reflect only a desire to improve timing performance and not a desire to guarantee task deadlines.

To support all three features, this research program proposes that the system maintain a *processing context* for each interpretation. In general, the context would list all relevant assumptions made by the problem solving processes at the time an interpretation was made or a piece of evidence produced. Specifically, the processing context would contain:

1. the parameter context: the values the front-end signal processing parameters (e.g., FFT-SIZE, PEAK-ENERGY-THRESHOLD, etc) had at the time the evidence for the interpretation was produced.

2. problem-solving assumptions about the signal. For example, which are the most critical sources to identify, what distortions like poor frequency resolution can be expected, etc.

3. environmental assumptions. For example, the number of sources considered active in the current scenario, the signal-to-noise ratio (e.g., the power ratio between those sources considered the focus of attention and all other identified and unidentified sources), etc.

4. the problem-solving goals in effect when the evidence and interpretations were produced. For example, the goal of reducing uncertainty resulting from alternative interpretations for the same data, or the goal of finding evidence for a particular important source.

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

6. the SPAs used to produce the evidence while the context was in effect and the time periods in which they were used.

For example, when a series of frequency spectra are generated by an FFT algorithm set for 512 points (and no changes for other parameter values), we say the series was generated under one parameter context with the parameter **FFT-SIZE** set to 512, and those spectra should refer to that context. Any microstream or other interpretations made with these spectra should also refer to that context structure. If the system had the assumption that no source was present whose frequency components could not be resolved by the FFT-SIZE value, that fact would be represented in the processing context, and the goals which were driving the generation of those spectra would also be included in the processing context.

In [27] the *multiple-views synthesis* capability is described as a desirable feature for next-generation signal interpretation systems because it would permit a system to fashion interpretations from evidence obtained from more than one parameter context. It is often the case that sufficient support evidence for a hypothesis can only be gathered by viewing the raw data under several parameter contexts, each of which precludes the availability of the others' evidence. Consider, for example, a pair of synchronized microstreams which have short attack regions (say, 0.1 sec) and are separated in frequency by 15 Hz. Assuming that the data was sampled at 10KHz, a series of FFTs with analysis 1024-point analysis windows applied to the data will provide adequate frequency resolution to detect both microstreams. The short attack period's energy variation will not be detectable, however, because the FFT windows are too wide and "blend" the energy of the 1000 points associated with the attack with that of the surrounding points. A series of FFTs with shorter analysis windows (say 256-points) would provide us with sufficient evidence to conclude the attack region was of the specified length, but it would not detect both microstreams. The shorter FFT windows would merge the two microstreams into one.

Only by processing the data under the two mutually-exclusive parameter contexts can all evidence for the two microstreams' descriptions be found. In the example it was seen that two properties (attack duration and frequency components) of a hypothesis required two different views of the same time region in the raw data. It is also possible that *different* time regions covered by a hypothesis may each require multiple views of the data from multiple reprocessings. These observations indicate that although the support evidence for a hypothesis may come from several contexts and should be so identified, for the sake of simplifying the generation of still higher-level hypotheses one would want to represent the intermediate-level hypothesis not

only as a composite but also as a single-context entity. Section 4.3 partially describes the design of a proposed context-mapping KS for generating the unified view of a multi-context hypothesis, and also discusses the use of this KS for mapping whole entities across processing context switches caused by global parameter adaptation.

In addition to simplifying the access procedure when searching for evidence obtained under different or related contexts, the processing context would provide a record of how portions of the signal were processed, how often they were reprocessed, how goals were satisfied or left unsatisfied while that portion was analyzed, and how environmental assumptions changed or were violated over time. This processing-context history would augment the IPUS SOU framework as a source of information describing *why* interpretations are uncertain and thus would provide a formal basis for generating reprocessing strategies to resolve discrepancy-related interpretation uncertainties[7]. Although beyond the immediate scope of this research program, the observation should be made that this information can also be useful in creating training instances to help the system learn reprocessing strategies or patterns for various scenarios, which in turn could help reduce the system's time spent in repeating spurious reprocessings when the most useful parameter context can be retrieved immediately.

Another new set of information planned for the system is a representation of the *costs* of different processing context changes. That is, the system needs a kind of state-transition graph, each of whose states represents a context equivalence class and whose transitions are labelled with costs associated with switching from the state representing the current front-end processing parameter values to a new context. A context equivalence class represents the set of all processing contexts which provide information at the same cost. "Cost" can include information about time requirements, classes of signals for which the context is and is not appropriate, etc. This global, generic information could be used by IPUS as a balance against the need to switch parameter contexts in order to obtain evidence for a specific source in the current scenario. It would represent context-independent assumptions about time and data. These costs would not only be useful from the obvious timing perspective, but also from a planning perspective. If two contexts are recommended, for example, the system could choose the one which provides not only the desired evidence, but also as widely-applicable as possible analysis for other sources which aren't the focus of reprocessing. Work being done in the *SPA model variety problem* (see [12, 27] and section 5 for more detail) indicates that the construction of such a graph is possible and practicable: [27] describes how the knowledge that an SPA is inappropriate for a small set A of signals can be extended to cover all signal sets U

where $A \subseteq U$.

The global parameter adaptation and source-model synthesis KSs are proposed in order to formalize control over selective processing streams. The global parameter adaptation KS would play two major roles in processing stream management. The first would be to select the initial front-end SPA set and initialize all the SPA parameters based on the available source database and any *a priori* environmental assumptions loaded into the testbed at start time. The second role would be to decide when to reconfigure the front-end SPA set in response to an inordinate amount of reprocessing or changes in the testbed's high-level processing goals. As aids in the configuration decisions, this KS must have access to the "context-transition graph" information as well as processing context information describing the relative identification accuracy required for each source and the possible discrepancies to which prospective contexts may give rise.

The proposed source-model synthesis KS's addition stems from the recognition that often source groupings will occur in complex environments such that the frequency or energy interactions among them cause the group to resemble not a simple superposition but a new model representing a complex merging of the group's isolated models. Such situations might give rise to excessive discrepancy-detection and reprocessing rates if separate models are maintained for each source, since it is possible for interactions among the group's sources to unpredictably mask or amplify features of the individual source models. The question therefore arises as to whether it is more efficient to represent the group by a dynamically-generated source-group model whose stream behaviors incorporate the wide range of frequency or energy variability in the group. The new model's looser constraints would reduce the number of discrepancies associated with the group's frequency and time regions, and would effectively represent an extension of the reprocessing concept from a strictly numeric-SPA basis to a higher-level model-reformulation basis. The source-model synthesis KS's role would be to decide when and how to create such models. This KS represents an attempt to integrate information from environmental assumptions, discrepancy detection and reprocessing to control the generation of new processing streams.

4.1 Integrating Selective Processing Streams into IPUS

4.1.1 Reprocessing KS

The component most closely related to IPUS's reprocessing capability is the Reprocessing KS itself. Many questions about the specification of this KS remain to be resolved. The KS currently views reprocessing to resolve uncertainty as a procedure which iteratively adjusts parameter values as specified by a script-like plan until the desired evidence is obtained. Right now the results of intermediate reprocessings are deleted from the blackboard with each new iteration. Certainly at least **some** of this data should be retained in the event that later circumstances require it. However, it seems wasteful of space to retain *every* intermediate reprocessing result. There should be some context-dependent mechanism for deciding how much intermediate information to retain. Perhaps the *purpose* or goals, of the current reprocessing should play a role in this to-be-developed mechanism. That is, reprocessing for finding highly-specific evidence while in the "discrepancy-detection, diagnosis, reprocessing" loop might be a good risk for deletion of all intermediate results while reprocessing for finding evidence in an aggregational problem solving mode might produce intermediate results with greater potential for future use because of their less-focused direction.

Another question involves the relationship between the concept of reprocessing and *levels of abstraction*. The role of reprocessing in the testbed's numeric-level processing and interpretation cycle seems reasonably well developed. The question is what role can the reprocessing concept play at higher abstraction levels. There appear to be at least two applications for reprocessing here: high-level interpretation reprocessing and source-model synthesis. High-level interpretation reprocessing pertains to the signal interpretations at or slightly below the source level. Does it seem possible to reprocess interpretations at the source level **without** having to reprocess all the lower-level data and then drive new evidence through the interpretation hierarchy, for instance? The preliminary answer appears to be "yes", but the kinds of parameterized algorithms which will fall under the sway of this higher-level reprocessing are not traditional signal processing algorithms. These algorithms will be implementations of psychoperceptual theories such as streaming in acoustic signals and illusory contouring in visual signals [5]. Work must be done in this area to create algorithms which have a unified formal theory similar in structure to that of traditional signal processing algorithms. Without such a theory, discrepancy detection and diagnosis at the higher levels (e.g. above the microstream level) would

be too ad hoc and would not provide a well-organized framework for controlling the reprocessing.

Source-model synthesis is a high level application of reprocessing which would be useful for scenarios with interacting (overlapping) sources. The current signal-source knowledgebase in IPUS contains models of *isolated* sources. This set of knowledge would not be adequate for source groupings where the frequency or energy interactions among them cause the group to resemble not a simple superposition but a new model representing a complex merging of the group's isolated models. Figure 12 provides a real-world example of this kind of situation. The role of a KS for reprocessing via source-model synthesis in this situation would be to reconfigure isolated models (or create new ones) which reflect the effects of these source interactions. A source-model synthesis KS would need access to such knowledge as the isolated source models' expected durations and the context-cost graph to determine the tradeoff between spending effort to build the group model and the estimated discrepancy rate expected for the duration of the source grouping. In addition to integrating information from environmental assumptions, discrepancy detection and reprocessing to control the generation of new processing streams, the formalization of this KS would represent an attempted answer to the general problem of modeling acoustic source interaction. In other acoustic domains such as speech recognition, this problem manifests itself in simultaneous multiple talker scenarios [32] and in the *coarticulation effect* [24][6] of connected speech.

It is true that the standard numeric-level approach to reprocessing via selective processing streams might work here and thus eliminate the need for model synthesis. One could conceivably filter out sources to find evidence for isolated models, and simply reprocess data for that evidence. This method is not only time-consuming, but also is not universally applicable. If one tries to filter raw data from a pair of similar-frequency sources so that one source is removed, it is highly likely that *both* sources would be removed from the analyzed spectrum. From a timing perspective, it is cheaper to reprocess the models to be matched than it is to reprocess the raw data in these cases.

The theme of processing streams' timing implications in this section indicates that this capability ought to be viewed from a control perspective not simply as a KS added to a signal understanding system, but as a new resource whose properties and costs must be weighed carefully before being requested. As an example of this,

[6]This refers to the blending of word boundaries so that the spectral signatures of words in connected speech appear different from their signatures when uttered in isolation.

[22] describes an IPUS test scenario where two alternative sources are supported by the same evidence and reprocessing must be engaged to find evidence which will differentiate between them. A parameter-adjustment KS proposes two parameter contexts for reprocessing which should provide positive evidence for one source and negative evidence for the alternative. One context requires decreasing the peak-detection energy threshold, whereas the other requires increasing the FFT size. From a local point of view, the first context is more desirable because looking at cached output from the original FFT algorithm with a lower threshold requires less time than reprocessing raw data with a larger FFT size. From a more global control perspective, however, the second context is preferable because its results are more discriminatory between the sources and will definitely give positive evidence to *one* of them, whereas the first context's results would at best find negative support for one source and no new support for the second. Examples like this reinforce the observation that IPUS needs a formal representation of reprocessing and/or context-switching costs.

Although mentioned earlier, it should be pointed out in this section for completeness that the benefits of learning "reprocessing macros" via examination of parameter contexts and number of reprocessings for given scenarios would be impressive. It seems highly plausible that the Reprocessing KS could be made to improve its reprocessing plans through learning larger/smaller parameter increments for new plans or composing new plans from old ones.

4.1.2 Diagnosis KS

The presence of multiple versions of the same data provided by reprocessing capabilities will require the design of the Diagnosis KS to be altered. A critical issue raised for this KS is how to extend it to perform diagnosis on entities constructed with supporting interpretations from different parameter contexts. This is important because most of the operators' preconditions are defined in terms of parameter values for *entire* entities. That is, the KS assumes that the whole time-frequency-energy region under consideration for diagnosis contains entities (e.g. contours and microstreams) all produced within the same parameter context, *and* that each entity does not represent a composite of parameter contexts. This assumption is made because the operators were designed to be applied *across the entire region under diagnosis*, not to particular entities.

A subproblem involves the question of how the search for distortion operators can be formally defined to take advantage of the presence of earlier reprocessings

(and diagnostic explanations in their accompanying problem-solving assumptions) of the current data being diagnosed for discrepancies. It seems plausible to constrain a distortion operator's applicability by an examination of reprocessings that were motivated by explanations containing that operator. If none of those reprocessing results were successful, the KS might need to eliminate the operator from further consideration.

A separate problem that must be addressed is the relationship between diagnosis of discrepancies at low levels (i.e. spectrum) and at "middle" levels (i.e. microstream). The term "middle" is used to distinguish this diagnosis from diagnosis at the source level, which probably will be different in that it will not be an extension of the middle level further along in time. In some cases, it is possible that a series of low-level discrepancies will also manifest themselves over time as a middle-level discrepancy, unless IPUS acts to resolve them as they appear. As an example, a series of "time-domain/STFT energy mismatch"[7] discrepancies could cause the appearance of a single "missing microstream" discrepancy at the middle level. It may be to the system's advantage to allow low-level discrepancies to accumulate and cause a middle-level discrepancy which, when resolved, will involve reprocessing which also eliminates the low-level discrepancies. In other situations, it may be more useful to handle the low-level discrepancies as they arise. The reprocessing capability can affect how this relationship is defined in that IPUS can use a comparison between the estimated reprocessing cost of resolving the series of low-level discrepancies and that of resolving the middle-level discrepancies as one criterion for deciding which level of diagnosis to pursue.

4.1.3 Discrepancy Detection KS

Continuing with the previous section remarks on selecting appropriate levels of diagnosis, we should note that discrepancy detection algorithms which are themselves parameterized for tolerance would be very useful here. If the low-level discrepancy detection's tolerance could be "squelched" or "amplified" in a systematic manner, this would be one way to implement diagnosis-level decisions. By raising the low-level discrepancy tolerance, for instance, the system would put into effect the decision that only medium-level discrepancies are to be produced and diagnosed.

A more important issue to be addressed for the Discrepancy Detection KS in the presence of multiple processing views concerns comparisons across parameter

[7] the signal energy in the time domain is appreciably greater than in the STFT spectrum

contexts. Given that several processing views are available, is it possible (and desirable) to detect discrepancies not just between signal processing algorithms' outputs, outputs and expectations, and outputs and *a priori* constraints, but between reprocessings? Currently this does not seem possible to the extent of justifying a new class of discrepancy, but it will probably be useful for the Discrepancy Detection KS to have access to earlier reprocessing results. Such information could be used to detect "discrepancy, reprocessing" patterns which could be learned for commonly-occurring sources. These patterns would let the system modify its expectations in the same way that we have the Diagnosis KS modify expectations in future blocks to include a description of what support evidence could look like under the parameter contexts which generated discrepancies in the current block.

4.1.4 Front-End Signal Processing Reconfiguration

Although a reprocessing capability is a powerful enhancement to signal understanding systems, it is still a time-consumptive feature. In section 4.1.1 we discussed using factors such as cost/benefits ratios to control the reprocessing invocation rate. Another means of reducing reprocessing invocations lies in selecting an initial front-end signal processing parameter context that is suited to the expected scenario. That is, knowledge about factors such as

- *source criticality:* which sources in a given environment are important and **must** be detected as quickly as possible versus which are less critical and may even be undetected without adverse effect;
- *expected source occurrence:* how likely are various sources to occur in a given environment;
- *source sequencing:* what sets of sources follow one another in a script-like fashion;
- *source confusability:* how difficult it is to distinguish critical sources from among all the currently identified environmental sources;
- *environmental interference:* what environmental features such as signals from unidentifiable sources, the state of the environment (see below), etc., can interfere with the detection and tracking of critical sources;

could be used to automatically select an initial parameter context that will "capture" a reasonable amount of evidence for as many sources in an environment as possible. The intellectual question to be addressed here is how to formally incorporate this and other information into a search paradigm whose performance can be systematically evaluated.

Beyond the ability to intelligently initialize itself, the IPUS system must be able to reconfigure its front-end processing parameter context periodically. In highly dynamic and signal-rich environments, no single parameter context is likely to minimize reprocessing invocations for very long. One could say that signal-generating environments undergo transitions between different phases. For example, a street corner may have a "busy phase" where high-amplitude car horns occur very commonly, followed by a "quiet phase" in the evening where only car engine sounds are infrequently detected. This example should not imply that these phases are necessarily predictable, only that they are easily differentiated. An IPUS-based system should be able to monitor the amount and kind of reprocessing it is performing to detect whether the current front-end processing is suited to the environment's current phase. If, for example, a large number of reprocessing invocations involving higher energy thresholds occurs, the system may need to reconfigure the front-end processing context to incorporate that new threshold value permanently.

Clearly, IPUS needs a KS for generating and applying models of the high-level environmental factors listed here, since the body of such knowledge is large and the interactions/tradeoffs inherent in this information is complex. This reconfiguration is termed *Global Parameter Adaptation* to distinguish it from the *Local Parameter Adaptation* that occurs during a single reprocessing invocation. Global parameter adaptation need not be performed only in response to the frequency and type of local parameter adaptations, either. It seems reasonable to have this process also controlled by time and space costs of the current front-end parameter context. Thus, even if the number of reprocessings is very low, we may still want to change the front-end context if the time requirements of, say, the FFT-size parameter value are causing the system to "slip behind" the data inflow rate.

4.2 Integrating Processing Assumptions with IPUS

The processing context is conceived primarily as a tool to support the consistent representation of processing assumptions and their uniform integration with all existing and planned IPUS components. "Processing assumptions" in general can include much information, ranging from the formal correctness of the SPAs' code to

the alleged completeness of the source model library (an unprovable proposition, indeed). Rather than attempt to model the entire universe in which the IPUS testbed exists and face the challenge of solving the *frame problem* [26] in all its generality, we will only consider the following limited set of assumptions for first-class representation. The major criterion for list inclusion was that the information represent facts subject to revision **in the presence of detected or inferred events in the monitored environment**.

1. Conditions for determining when a source's identification is sufficient; that is, the criteria for terminating the evidential search for a source hypothesis. These include not only numeric evidence credibility thresholds but also high-level information such as observed critical source regions[8].

2. Classes of sources assumed active in the scenario. Such knowledge can influence the choice of bottom-up streaming strategies.

3. What evidence is considered distorted, but acceptable.

4. Observed discrepancies, their related interpretation hypotheses, and explanations (if any).

5. Reprocessing strategies performed and their motivating discrepancy explanations.

6. Expected distortions given SPA parameters and current environmental sources.

4.2.1 Discrepancy Detection

Information about what classes of sources are current active might be useful as a parameter governing how sensitive discrepancy detection processes must be in classifying signal behavior under different analytic views as discrepancies. However, an alternative approach toward integrating these assumptions might be to include them in focusing heuristic function definitions for choosing what discrepancies to pursue next and keep the discrepancy detection mechanism at a single level of sensitivity. Which approach should be taken will depend on the design (and knowledge-engineering) modularity each provides.

[8]For example, in the presence of competing sounds, one might be satisfied that the sound of a bell being struck was detected even though evidence was only detected for the bell's "gong" and none was found for the bell's decaying vibrations.

4.2.2 Discrepancy Diagnosis

Adding a parameter to this knowledge source describing available time and desired explanation certainty should provide a clean mechanism for expressing search and data approximation time-based constraints. The time constraints in the current processing context should be used by control plans to generate a value for this KS parameter. On the basis of this parameter, the KS would decide how deep in the data abstraction hierarchy it should perform explanation verification. How the control plans will decide appropriate constraints, and what the exact time requirements and certainty for each abstraction level are remain open questions.

The presence of environmental assumptions must be integrated with the discrepancy diagnosis KS to enable their use in more sophisticated diagnostic strategies for determining why a signal expectation was violated and whether or not to reprocess data to find new corroborating data for the explanation. If, for example, the new IPUS testbed has access to the assumption that there are no new sound sources that have appeared in the environment, then the discrepancy diagnosis KS should be able to eliminate the possibility of a newly-occurring source masking a previously-observed source as an explanation for any "fadeout" or loss of signal detected in the observed source, and instead focus attention on distortion operators that don't depend on the existence of new sources.

This KS, as well as the Reprocessing KS, will play major roles in a belief revision mechanism for updating or retracting environmental and problem-solving assumptions. If a diagnosis cannot be found given a set of environmental assumptions, or if a reprocessing strategy plan set is exhausted without producing desired evidence, then not only should negative evidence be attributed to the existence of the desired data, but the environmental and signal processing assumptions may also require annotation with negative evidence. The integration of this mechanism with RESUN's SOU framework is an important open problem.

A preliminary step has already been taken toward the tighter integration of discrepancy diagnosis with discrepancy detection and processing contexts [22]. It involves the modification of expectations for future support evidence's appearance or quality and takes the form of an extension to the discrepancy diagnosis KS that provides each distortion operator with a distortion specification (represented as a logical implication of the form *IF operator-preconds THEN distortion-pattern*) of how expected data can appear distorted under processing parameters' current values. When an explanatory operator sequence is found, all operators' preconditions and distortion specifications and are conjunctively combined to form a support spec-

ification for the hypotheses involved in the original discrepancy. The support specification has the form *IF (precond*$_1$ *AND ... AND precond*$_n$*) THEN (distortion*$_1$ *AND ... AND distortion*$_n$*)*. The specification's distortion pattern locally modifies the high quality normally required of all evidence for consideration as support for an expectation and permits the use of distorted evidence (*without* raising a discrepancy) for extending the annotated hypothesis as long as the processing context (e.g., the preconditions) that caused the current distortion persists. If the discrepancy explanation enabled the reprocessing KS to find a strategy to eliminate the discrepancy, the hypotheses involved in the discrepancy have their expectations annotated with the support specification; if no reprocessing strategy was found, the specification is discarded.

This extension has shown its utility by serving to reduce the amount of reprocessing performed by the testbed. Nevertheless, its reliance on simple conjunction for support specification generation requires refinement. The preconditions currently specify the precise context parameter values under which the distortion is expected to persist; the extension's generality would be improved if instead a context equivalence class could be generated as the specification's precondition.

4.2.3 Reprocessing

In the same vein as the idea of goal-based evidence deletion/retention mentioned in section 4.1.1 is the general concept of data maintenance. Every interpretation system will eventually exhaust its space for the interpretation database, and must incorporate some means of "forgetting" evidence and deleting it to provide room for storing future interpretations. Explicit environmental assumptions such as past noise levels and source importances might play useful roles in deciding what low-quality evidence to delete and what evidence to maintain. These factors might be useful as mitigating factors affecting a general maintenance mechanism which deletes evidence solely on the basis of aging[9]. Instead, evidence from important sources or highly-reprocessed time regions might be profitably preserved for longer times than the defaults specified for their abstraction levels. Work in this area will make use of existing techniques in applying decision theory [18] to the problem of allocating computational resources. However, these techniques require decision models, and the development of models relating high-level concepts such as source criticality and reprocessing redundancy to the decision of deleting or retaining data at various

[9]That is, evidence at abstraction level X should be deleted after it becomes N_X seconds old.

abstraction levels will represent new contributions to acoustic interpretation system design.

4.3 Integrating Multiple Views Synthesis with IPUS

Parameter contexts can provide a useful mechanism for combining multiple views, or data obtained under different processing parameters. Because each interpretation datum will be marked by the processing context under which it was created, it will be possible to use Fourier theory to *map* interpretations across contexts during the search for support evidence in reprocessed data.

As an example, consider a source that can be modeled by

$$s(t) = \cos(2\pi 1200t) + \cos(2\pi 1220t) + f(t) \tag{1}$$

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. This source 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 13). Assume at some time t an impulsive (approximately 0.1 sec duration) source appears, and is not detected by the STFT algorithm output, but time-domain tests such as average signal energy analysis indicate its possible presence, necessitating reprocessing in the higher-energy region by the STFT with a shorter (say 256 points) analysis window.

When the signal data was originally processed by an STFT algorithm with an analysis window length of 1024, an entire beat period was analyzed at a time, and the magnitude assigned to the two sinusoids in the STFT's output spectra was relatively steady. When the data is reprocessed with a 256-point analysis window, however, the window's data will only cover a quarter 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 the contouring method 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 unexpected appearance of 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), one can use Fourier theory to map these new contours' energy values and behaviors into ranges in the target context and, if these ranges fall in frequency-energy regions where support is desired, the 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.[10]

The choice of casting multiple views synthesis as a mapping between contexts raises several questions: how is the target context chosen; during discrepancy detection or reprocessing goal verification, must all entities in the source context overlapping the time region of interest be mapped back to the target context; when global parameter adaptation occurs, must all work in the previous front-end context be mapped into the new context; from a knowledge engineering perspective, how can all the mapping knowledge be modularly and succinctly expressed? The preliminary answer to the first question is that the target context should be the front-end parameter context selected by the global parameter adaptation KS. If only parameter contexts are considered (as opposed to the complete processing context), the preliminary answer to the second and third questions is "probably, NO." When entire processing contexts are considered (with their environmental assumptions), the problems of what and how to map become much more complicated. They are similar to that encountered in truth maintenance systems [10] when facts assumed in one possible world are disproven or assumed false in another possible world. To control this interaction across contexts in order to minimize mapping, we currently envisage a two-step indexing scheme for determining the applicable mapping knowledge:

1. mapping procedures are selected based on what parameter values are different between the source and target contexts

2. only those mappings whose environmental and problem-solving assumptions (e.g. the set of all significant frequency components is assumed to be known completely) are met by the source processing context are applied.

The fourth question seems to require a KS initialized with mapping knowledge (as

[10] As an aside, in this example there should be another application of context mapping in the reverse direction, where the energy of the impulsive source's contours under the reprocessing context would be mapped into lower energy contours in the original parameter context. This high energy results from the reprocessing context's smaller reprocessing analysis window covering only the high energy signal region produced by the impulse. A wider window "averages" the impulse energy with the surrounding lower energy signal, producing a lower observed energy for the impulse in the original context.

described in the previous example) defined on the crossproduct of the entire SPA set available to the perceptual system.

5 Research Evaluation Approaches

We expect the final product of this paper's work to provide several contributions to the field of perceptual system design:

1. A generalized architecture formally integrating the key IPUS processes of discrepancy detection, diagnosis, reprocessing, and differential diagnosis with explicit processing assumptions, selective processing streams, and multiple views synthesis.

2. An answer to the problem of how the multiple views afforded by selective processing streams can be exploited by problem-solving strategies augmented with explicit processing assumptions.

3. A demonstration of the architecture's applicability and generality.

4. A quantitative measure of the reprocessing capability's importance to signal interpretation quality.

5. A viable platform for future experimentation on psychoacoustic streaming theories [5] and control strategies for acoustic signal interpretation.

Given these goals, there are two perspectives from which this research program's results should be evaluated. The first is qualitative and concerns the organization provided by the new architecture features and the formalized interactions among the primary knowledge sources (goals 1, 2, and 3). The second perspective is quantitative in nature and concerns assessing the reprocessing capability's utility in interpretation quality and measuring the effectiveness of control strategies and focusing heuristics developed in the course of testing system components on real-world acoustic environments (goals 4 and 5).

The generalized architecture resulting from this research should be evaluated on several organizational bases:

- How localized is signal processing knowledge? Is it well separated from control knowledge and concentrated in knowledge sources? How easy is it to add new

knowledge about instances of discrepancy classes, distortions, or reprocessing plans?

- How application-independent is the architecture? Can the architecture flexibly accommodate additions to source model definitions or new classes of environmental assumptions?
- How cleanly have multiple views and explicit assumptions been integrated with all architecture components? Does a formal framework exist for specifying new application domains' interpretation-level knowledge sources in the presence of these new features?
- What kind of power is the reprocessing capability providing? Is this power unique to the availability of multiple views, or can it be attained through other means (e.g. in cleverly-crafted interpretation knowledge sources)?
- Is there a formal framework for estimating the amount of reprocessing that will be performed in an environment with a given set of sources?
- What kinds of environments can be monitored by the architecture? For what classes of monitoring tasks in these environments is the IPUS approach suited?

In connection with the last two points, the sound testbed's development thus far has had a strong empirical nature. That is, given knowledge of the scenarios that could be encountered, the designers supplied the testbed with a set of SPAs they believed adequate for the interpretation tasks it could face. However, no formal analysis was ever performed on the scenarios to determine a priori what algorithms would be needed and *where* in the data streams processing by two or more SPAs would be required. Work in this area would rely heavily on work being done at Boston University on the *SPA model variety problem* [12]. This problem focuses on the relationship between SPAs and the classes of signals for which they can produce undistorted outputs. A signal understanding system needs to use more than one SPA if there does not exist a single SPA that can produce undistorted output for all the possible input signals in the given application domain. In other words, the input signal must satisfy the conditions in the data-model for that SPA. If the SPA does not satisfy the conditions, it is said to suffer from a model variety problem with respect to the input signals. Formal analysis of this relationship between SPAs and input signal characteristics would permit one to predictably tailor the testbed's SPA database to specific scenario classes. This ability would also enable one to formally

evaluate the focusing heuristics and control plans effectiveness in terms of the ratio between the amount of data actually reprocessed and the estimated required amount of reprocessing.

With regard to quantitative analysis of the framework, two sets of experiments recommend themselves. The first deals with determining the reprocessing capability's importance to interpretation quality, and the second deals with evaluating acoustic signal interpretation strategies encoded in control plans and focusing heuristics.

In the first experiment set, an interpretation-quality baseline for the IPUS testbed would be established for given scenarios by running the system with the selective processing streams capability disabled. Essentially this would represent the quality of a fixed front-end processing system. SPA parameters would be initialized to the values necessary for "high-quality" (numeric definition to be decided) identification *and* tracking (determination of times of source onset and fadeout) of the most critical source in the scenario. The baseline would consist of the total run time and the weighted number of correct source identifications. This weighted number of correctly-identified sources is defined as

$$W = \sum_{i=1}^{N} b_i(d_i/D_i) - \sum_{j=1}^{M} b_j(d_j/D_j^e), \tag{2}$$

where N is the number of sources actually in the scenario and M is the number of false-alarm source identifications. The value b_i is the testbed's overall belief value associated with the i-th source, while d_i is the duration for which the i-th source was tracked in the scenario, D_i is the actual duration of the source in the scenario, and D_i^e is the average duration length for the i-th source in the source database. The same scenarios (and parameter initializations) would then be run several more times on the IPUS testbed, each time with increasing limits on the amount of reprocessing (in terms of data points) that can be performed, until finally an unlimited amount of reprocessing was permitted. In each case, the same measurements as those for the baseline would be taken.

If the reprocessing capability is truly useful, then a plot of the experiment results should show an initial increase in the weighted number of correct source identifications over the baseline as processing time increases. The plot should also show a "diminishing returns" effect as reprocessing limits grow. That is, a point would be reached where source-identification beliefs grow only slightly even with large amounts of reprocessing. Such an experiment suite would provide an answer to

the question of whether the overhead incurred by the processing context and the selective processing streams is justified by any significant increase in the testbed's interpretation quality and would satisfy evaluation goal 4.

In the course of formalizing the architecture, several sets of control plans and focusing heuristics will be developed to test different control strategies, some of which will be based on experimental psychoacoustic observations [5]. This control knowledge can be evaluated quantitatively in the second experiment set in terms of how often the strategy employs reprocessing over a wide range of scenarios versus source identification certainties in the presence of increasing interference between sources in a scenario. The second experiment suite would be based on the following: given a fixed set of K sound sources, generate a large set S of random source orderings. For each ordering $s_i \in S$, one source from the database is designated as the most important source to identify and track, while the others are given random, but lower, importance levels, indicating the lower degrees of effort expected for their detection. Set S would be copied M times, each copy's lower-importance sources having increasing energies relative to the most important source. The IPUS testbed current control strategy would be run on each scenario set, and the same statistics as in the first experiment set (testbed run time, amount of reprocessing, and weighted number of correct source identifications) would be collected for each scenario. A plot of the data for each strategy would indicate how often (and at what benefit in terms of source identification certainty) reprocessing is performed by the strategy as less-important sources' volume (e.g., "noise") increases.

As a side benefit, work in generating the test suites' scenarios should also provide some insight into the design of benchmarks in the acoustic signal interpretation domain. In [36] the problem of designing benchmarks for visual interpretation systems has been identified as a significant research problem for the machine perception research field.

6 Acknowledgments

This paper is based upon work supported by the Rome Air Development Center of the Air Force Systems Command under contract F30602-91-C-0038. The content of the information does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

I wish to thank my advisor Victor Lesser for his guidance and incredible patience while I was developing the ideas in this paper. I also owe a debt of gratitude to

Hamid Nawab and Erkan Dorken in the ECS department at Boston University for their signal processing technical advice.

References

[1] J. F. Allen and P. J. Hayes, "A Common-Sense Theory of Time", *IJCAI '85 Proceedings*, pp 528–531.

[2] B. Bell, L. F. Pau, "Context Knowledge and Search Control Issues in Object-Oriented Prolog-Based Image Understanding," *ECAI '90 Proceedings.*

[3] B. Bell, L. F. Pau, "Contour Tracking and Corner Detection in a Logic Programming Environment", *IEEE Transactions on Pattern Recognition and Machine Intelligence*, August 1990.

[4] N. Bitar, E. Dorken, D. Paneras and H. Nawab, "Integration of STFT and Wigner Analysis in a Knowledge-Based Sound Understanding System," *IEEE ICASSP '92 Proceedings*, March 1992.

[5] A. Bregman, "Auditory Scene Analysis: The Perceptual Organization of Sound," MIT Press, 1990.

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

[7] N. Carver, *Sophisticated Control for Interpretation: Planning to Resolve Sources of Uncertainty*, PhD Thesis, Computer Science Dept., University of Massachusetts, 1990.

[8] T. Claasen and W. Meclenbrauker, "The Wigner Distribution: A Tool for Time-Frequency Signal Analysis," *Phillips J. Res.*, vol. 35, pp. 276–350, 1980.

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

[10] J. deKleer, "An Assumption-Based Truth Maintenance System," *Artificial Intelligence*, vol. 29, pp. 241-288, 1986.

[11] R. De Mori, L. Lam, and M. Gilloux, "Learning and Plan Refinement in a Knowledge-Based System for Automatic Speech Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Feb 1987, pp 289-305.

[12] E. Dorken, H. Nawab, and V. Lesser, "Extended Model Variety Analysis for Integrated Processing and Understanding of Signals," *IEEE ICASSP '92 Proceedings*, March 1992.

[13] W. Dove, *Knowledge-Based Pitch Detection*, PhD Thesis, MIT, 1986.

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

[15] F. Giannesini, H. Kanoui, R. Pasero, and M. van Caneghem, *Prolog*, Reading, MA: Addison Wesley, 1986.

[16] G. Goertzel, "An Algorithm for the Evaluation of Finite Trigonometric Series," *American Mathematics Monthly*, Vol. 65, pp. 34–35, Jan 1958.

[17] B. Hayes-Roth, R. Washington, R. Hewett, M. Hewett, and A. Seiver. "Intelligent Monitoring and Control", *IJCAI '89 Proceedings*, pp. 243–249.

[18] E. J. Horvitz, "Reasoning Under Varying and Uncertain Resource Constraints," *AAAI '88 Proceedings*, pp. 111–116.

[19] E. Hudlická, and V. Lesser, "Meta-Level Control Through Fault Detection and Diagnosis," *AAAI '84 Proceedings*, 1984, pp. 153–161.

[20] C. Kohl, A. Hanson and E. Reisman, "A Goal-Directed Intermediate Level Executive for Image Interpretation," *IJCAI '87 Proceedings*, pp 811–814.

[21] W. A. Lea, "The Value of Speech Recognition Systems," Trends in Speech Recognition, ch. 1, Speech Science Publications, 1986.

[22] V. Lesser, H. Nawab, M. Bhandaru, Z. Cvetanović, E. Dorken, I. Gallastegi, and F. Klassner, *Integrated Signal Processing and Signal Understanding*, Technical Report 91-34, Computer Science Dept., University of Massachusetts, 1991.

[23] V. Lesser and D. Corkill, "The Distributed Vehicle Monitoring Testbed: A Tool for Investigating Distributed Problem Solving Networks," *AI Magazine*, vol. 4, no. 3, pp. 15-33, 1983.

[24] B. Lowerre, D. Reddy, "The HARPY Speech Understanding System," in *Trends in Speech Recognition*, Prentice-Hall, 1980; Speech Science Publications, 1986.

[25] Luo and Kay, "Multisensor Integration and Fusion in Intelligent Systems," *IEEE Transactions on Systems, Man and Cybernetics.* Sept/Oct 1989.

[26] J. McCarthy and P. Hayes, "Some Philosophical Problems from the Standpoint of Artificial Intelligence," in *Machine Intelligence 4*, Michie and Meltzer, editors. Edinburgh: Edinburgh University Press, 1969, pp. 463–502.

[27] H. Nawab and V. Lesser, "Integrated Processing and Understanding of Signals," ch 6, *Knowledge-Based Signal Processing*, A. Oppenheim and H. Nawab, editors, 1991.

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

[29] H. Nawab, V. Lesser, E. Milios, "Diagnosis Using the Underlying Theory of a Signal Processing System," *IEEE Transactions on Systems, Man and Cybernetics. Special Issue on Diagnostic Reasoning.* May/June 1987.

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

[31] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, Englewood Cliffs, NJ: Prentice Hall, 1989.

[32] T. W. Parsons, "Separation of Speech from Interfering Speech by Means of Harmonic Selection," *Journal of the Acoustic Society of America*, vol 60, no. 4, pp 911–918, Oct. 1976.

[33] Y. Peng and J. Reggia, "Plausibility of Diagnositic Hypotheses: The Nature of Simplicity," *AAAI '86 Proceedings*, pp 140–145.

[34] A. Rosenfeld, "Image Analysis: Problems, Progress, and Prospects," *Pattern Recognition*, (17)1:3-12, 1984.

[35] D. Seborg, et al, "Adaptive Control Strategies for Process Control: A Survey," *AIChE Journal*, vol. 32, no. 6, pp. 881-913, June 1986.

[36] M. Swain, and M. Stricker, eds. *Promising Directions in Active Vision,* NSF Active Vision Workshop, Technical Report CS 91-27, Computer Science Dept, University of Chicago, 1991.

[37] M. Williams, "Hierarchical Multi-Expert Signal Understanding," in *Blackboard Systems*, Robert Engelmore and Anthony Morgan, editors, Addison-Wesley, 1988.

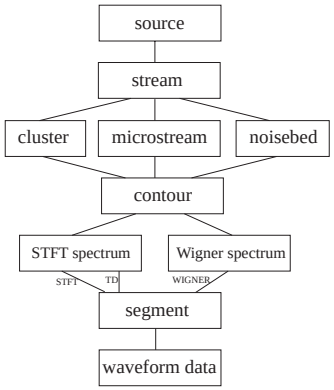

Figure 4: *Evidence abstraction hierarchy used in the IPUS testbed. Abstractions serve as support for those abstractions immediately above them.*

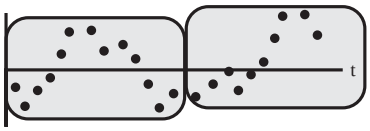

Figure 5: SEGMENT LEVEL: *A segment is a collection of raw data points for which such time-domain statistics such as zero-crossing density, average energy, etc, are maintained. Numeric-level SPAs operate on one segment at a time.*

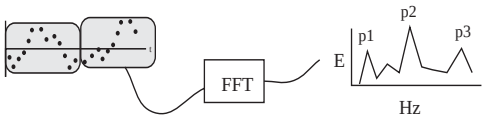

Figure 6: SPECTRUM LEVEL: *The second level consists of spectrum hypotheses derived for each waveform segment through Fourier-Transform-based algorithms such as the STFT and Wigner-Distribution [8] algorithms and peak-picking algorithms.*

Figure 9: NOISEBED LEVEL: *The fifth evidence abstraction level contains noisebed hypotheses supported by one or more contour clusters. Noisebeds represent the wideband component of a source's acoustic signature. Usually microstreams form "ridges" on top of noisebed "plateaux", but not every noisebed has an associated microstream.*

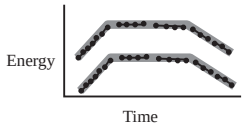

Figure 10: STREAM LEVEL: *Groups of microstreams and/or noisebeds synchronized according to time and/or some psychoacoustic criteria (e.g., harmonic sets, frequency separation) support stream hypotheses in the sixth level.*

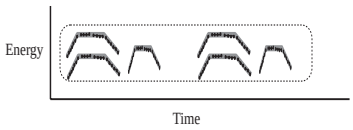

Figure 11: SOURCE LEVEL: *At the seventh level, sequences of stream hypotheses are used to support sound-source hypotheses.*

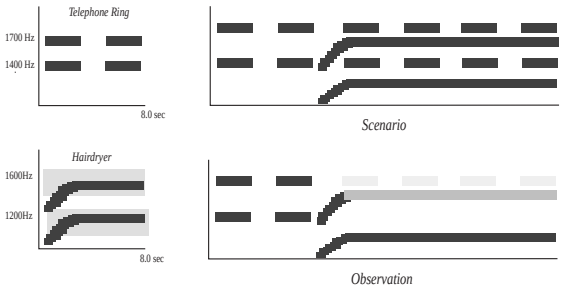

Figure 12: *An illustration of source interaction that could require source-model synthesis. The two left-hand graphs show isolated source models for a telephone ring (top) and a hairdryer fan (bottom). The top right-hand figure shows the conceptual appearance of the scenario being analyzed (simple superposition) and the bottom right-hand figure shows what is actually observed. In this case the telephone rings' lower-frequency microstreams are totally masked while their higher-frequency microstreams are significantly masked by the noisebeds of the hairdryer. It is more expensive to reprocess the data with filters to find the evidence for isolated sources than it is to construct a source-model which combines both sources and use available data for its support.*

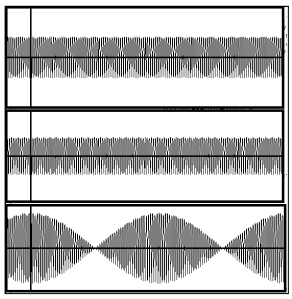

Figure 13: *A simple illustration of the beat phenomenon. The top two graphs indicate pure cosine waves at 1200 and 1220 Hz, respectively. The bottom graph shows the sum of the two cosine waves. Note the induced 10 Hz beat in the waves. All three graphs span 1100 data points, or 0.11 seconds. The third graph shows slightly more than one beat period.*