Intelligent Analysis of Composite Acoustic Signals

by: John M. Chowning

The "proposed" research concerns the architecture of an intelligent system for analyzing composite acoustic signals in the music domain.

The proposed methodology attempts to integrate knowledge from signal processing, perception, structural analysis of signals, reasoning for signal understanding and learning. The *scientific quality* of the proposal is very high. There are very few research groups in the world that could carry on research like this. The applicants seem to have excellent ideas on how to make use of their knowledge in different domains. The *importance* of the research seems very high to being the complex analysis of complex acoustic signals, a very important problem especially when tracking of simultaneous acoustic signals is involved.

The *capability* of the applicants seems to be very high. Their analysis of the problem is deep on all aspects. They seem to have an excellent perception on AI and signal processing techniques that could be useful for the problem they wish to solve. They should have, perhaps insisted more on the difficulties of learning rules automatically, being Machine Learning a discipline still far from producing solutions to real complex problems.

The only aspect that is not particularly exciting in this proposal is the *publication record* of the applicants describing work supported by previous NSF Contracts.

I am not familiar with the amount of publications produced by the Computer Music Community. I was a little surprised not to see any journal publication in the list on page 2. I must say that publications listed on page 2 appeared in the Proceedings of important conferences. Having a paper accepted at IJCAI-85 must have been particularly difficult for a researcher in computer music.

The budget seems reasonable to me.

In conclusion, I consider the proposal content excellent, but I am rating the overall proposal "very good" because I would have preferred seeing more evidence of previous work disseminated in International Journals.

Review: "Intelligent Analysis of Composite Acoustic Signals"
NSF Proposal Num: 86-13574

This is a proposal to continue development of a system for acoustic analysis of complex audio signals – music or otherwise – using pattern recognition techniques that analyze a signal both in frequency and time to separate and recognize repeated structures. It includes proposals for novel signal processing techniques as well as AI architectures to achieve this end. This is a continuation of a project begun in 1983 on which they seem to have made considerable progress. Very briefly, I strongly recommend support of this project. The problem is an interesting and important one and they clearly have the skills to make excellent progess on it. This work contributes to the development of systems capable of synthetic perception.

In their previous work, the authors have tackled some very difficult problems with considerable success, as far as one can tell from their publications. It is difficult to tell exactly since they tend to illustrate each system with only a single example. But they appear to have systems that distinguish the edges of a string of lagato notes played on a cello, recognize rhythmic figures in audio signals and resolve the separate notes in a simultaneously played piano chord. These are considerable achievements.

One major difference between the proposed project and the earlier one is that they now are interested in building a more general auditory perception system – one not restricted to music. I am of two minds as to whether this is a good move. On one hand, it makes sense since auditory stream segregation is a more basic and primitive skill than music perception. Presumably even fairly primitive nervous systems must do a pretty good job of this (flies and fish probably do it OK, but cats are probably as good at this as humans). Probably only people really perceive musical structure. Their proposed new problem domain (although they do not commit themselves to it wholeheartedly) is the perception of multiple birdsongs, rather like the sound background in a Tarzan movie, one supposes. This is a good problem since it allows simulation of the "cocktail party effect" but with fairly simple signals to analyze.

I only hope they record their *own* birdsong tapes for starters since field recordings will be *very* noisy raising the difficult question of deciding whether or not there is a signal there at all along with the problem of what the pattern is when do you know it's there. In fact, the signal detection question could really only be solved *after* the problem of pattern detection is solved. That is, listeners ability to detect patterns in noise is better than one would predict from their sensitivity to random unpredictable signals in noise. This justifies using clean signals in early stages of this work. And, of course, their reasoning may be (although it does not seem to be spelled out) that to differentiate musical voices they must first solve the auditory segregation problem.

A second advantage of the multiple birdsong task is that the notes of many birdsongs are nearly sinusoidal (although third-harmonic distortion can easily occur during analog processing). Real birdsongs, however, are not as regularly structured as music and seem to be inherently probabilistic.

Still, much simpler problems might be ones that simulate the ecological auditory streaming problem faced by typical land animals: simultaneous chair squeeks, footfalls, a mesquito, bus driving by, a kid on an intermittently squeeky swing, etc. It just might be that these would be easier yet still challenging.

On the other hand, there are reasons why I regret seeing them abandon the focus on music. Music offers the opportunity to work on integrating the rhythm detection system (which they have begun) and the melodic pattern detector (also begun). Since melodies change notes at points in the rhythm cycle that are metrically strong, ie, at points that are small rational-fractions of the cycle. These two systems should be brought together. A blackboard may be a reasonable way to achieve this integration – but I think the blackboard should, for music, move quickly from a flat one to a cylindrical or toroidal one. For birdsong, the blackboard will have to be flat.

I applaud their increased attention to the psychological literature on perceptual streaming (especially that of Bregman and his group). It is very likely that results from research on human performance will be useful. In this proposal, however, they are not very explicit on how this work is relevant and what use will be made of it. In fact, I sense an incipient contradiction between their endorsement of the Bounded-Q signal processing, basically a signal-processing hack that improves the resolution of the upper frequencies beyond what a normal FFT would produce, and the interest in building auditory models directly into their system. There is nothing remotely analogous to this in human auditory function, for which the a log-scaled FFT is a reasonable first-order approximation. It is likely that human auditory systems achieve their streaming and frequency discrimination skills through techniques having no resemblance to this trick – possibly via time-domain techniques. (Obviously, this not a reason not to do it; I am only noting what appears to be a discrepancy between the two approaches within the proposal. This difference of viewpoint may reflect differences of orientation within the research group.)

Basically, I think they have a very promising idea and address an extremely important problem in the development of intelligent systems that approach human performance. They have a good chance at success given their previous record of productivity. They should definitely be supported.

NATIONAL SCIENCE FOUNDATION

PROPOSAL EVALUATION FORM

NSF Form 1 (4/84)
Supersedes All Previous Editions

PROPOSAL NO.	INSTITUTION	PLEASE RETURN BY
DCR-8613574	Stanford Univ	07/19/86

PRINCIPAL INVESTIGATOR	NSF PROGRAM
John M. Chowning	INTELLIGENT SYSTEMS PROGR

TITLE
Intelligent Analysis of Composite Acoustic Signals

Please evaluate this proposal using the criteria presented on the back of this review form. Continue on additional sheet(s) as necessary. In a broad sense, the goals of this proposal are well thought out and worth supporting. Furthermore, the applicants are clearly very well qualified to carry out the research. Consequently, I would recommend support of the research.

The project is aimed at development of a system for assisting or automatically carrying out the analysis of acoustic sounds, combining front-end signal processing with AI technology. The proposal articulates some broad philosophies and concepts but is not particularly detailed in how to accomplish the next phase of the system. Consequently, the positive recommendation is based primarily on the sense that they have generally chosen a project with good potential, and that the investigators are sophisticated and creative. In particular, I believe that their choice of automatic analysis of bird songs is a good one and that results obtained will be more generally applicable, perhaps even to analysis of human speech in a multispeaker environment.

Apparently, an important part of their proposed program involves the front end signal processing. I'm personally not convinced that their Bounded-Q Frequency Transform is as clever or important as they suggest. In referring to the "surprising efficiency" of the method, they do not account for the additional computation required for the lowpass filtering required prior to each downsampling stage. Furthermore, while the procedure does give multiple analyses with bandwidth and time resolution differences of successive factors of two, this is highly redundant (as they recognize). Furthermore, the basic procedure is restricted to time and bandwidth resolution jumps of integer amounts if much more sophisticated decimation/interpolation schemes aren't used.

OVERALL RATING: ☐ EXCELLENT ☒ VERY GOOD ☐ GOOD ☐ FAIR ☐ POOR

NATIONAL SCIENCE FOUNDATION

PROPOSAL EVALUATION FORM

NSF Form 1 (4/84)
Supersedes All Previous Editions

PROPOSAL NO.	INSTITUTION	PLEASE RETURN BY
DCR-8613574	Stanford Univ	07/19/86

PRINCIPAL INVESTIGATOR	NSF PROGRAM
John M. Chowning	INTELLIGENT SYSTEMS PROGR

TITLE: Intelligent Analysis of Composite Acoustic Signals

Please evaluate this proposal using the criteria presented on the back of this review form. Continue on additional sheet(s) as necessary.

SChowning and Mont-Reynaud have proposed a very promising attack on an interesting and important problem. The three main investigators, Chowning, Mont-Reynaud, and Chafe have already established leading reputations in the required technologies - artificial intelligence signal processing and computer music. The funds requested seem reasonable and well matched to the size of the effort. Most of the computing hardware and software is in place and has already demonstrated productivity. The proposal is excellent in all aspects.

The problem of automatic analysis of multi-voiced music and other acoustic signals is theoretically interesting but in addition a system to automatically transcribe and write out a score for a musical performance would be of great value in understanding what constitutes good music performance. The question is theoretically attractive because it is a recognition question related, for example, to speech recognition. However, music is enough simpler than speech so that faster progress should be possible. The experimenters are approaching music recognition in a sufficiently general way so their results may also apply to other recognition domains, and in fact they plan to try another domain - bird sounds.

The proposed attack consists of a combination of signal processing analysis plus AI learning. Such a double discipline approach is essential in that there is ample past work to show that neither signal processing or AI by itself gets very far.

The BQFT transform is a very attractive way to analyze the raw signals. It is computationally efficient, it is easy to trade off time resolution against frequency resolution, and the general idea of a "constant Q" analysis is intuitively sensible.

I strongly recommend support of this fine proposal. The work is likely to lead to results of both immediate and long range importance. Society will get good value for its money.

OVERALL RATING: [X] EXCELLENT [ ] VERY GOOD [ ] GOOD [ ] FAIR [ ] POOR

NATIONAL SCIENCE FOUNDATION

PROPOSAL EVALUATION FORM

NSF Form 1 (4/84)
Supersedes All Previous Editions

PROPOSAL NO.	INSTITUTION	PLEASE RETURN BY
DCR-8613574	Stanford Univ	07/19/86

PRINCIPAL INVESTIGATOR	NSF PROGRAM
John M. Chowning	INTELLIGENT SYSTEMS PROGR

TITLE: Intelligent Analysis of Composite Acoustic Signals

Please evaluate this proposal using the criteria presented on the back of this review form. Continue on additional sheet(s) as necessary.

Chowning's Center for Computer Research in Music and Acoustics (CCRMA) at Stanford is, along with the similar center in Paris (IRCAM), one or the two or three leading centers in the world for computer-based research on musical acoustics. Chowning's Center continues to attract and to train first-rate young researchers in the field, and to generate important research and highly influential musical and technological developments. The present proposal clearly reflects the intention of Chowning and his group to maintain CCRMA at the forefront of research in acoustics and auditory percepion as those disciplines relate, particularly, to the ways in which humans perceptually parse the continuous time varying pressure waves at the two ears into discrete, spatially localized external sources (whether natual, human, or musical), each with its own unique characteristics. I believe that the National Science Foundation should ensure that this work will continue to go forward.

OVERALL RATING: ☑ EXCELLENT ☐ VERY GOOD ☐ GOOD ☐ FAIR ☐ POOR

NATIONAL SCIENCE FOUNDATION

PROPOSAL EVALUATION FORM

NSF Form 1 (4/84)
Supersedes All Previous Editions

PROPOSAL NO.	INSTITUTION	PLEASE RETURN BY
DCR-8613574	Stanford Univ	07/19/86

PRINCIPAL INVESTIGATOR	NSF PROGRAM
John M. Chowning	INTELLIGENT SYSTEMS PROGR

TITLE: Intelligent Analysis of Composite Acoustic Signals

"Please evaluate this proposal using the criteria presented on the back of this review form. Continue on additional sheet(s) as necessary.

This is an impressive scientific proposal by a group of top-notch researchers. The work touches upon many disciplines and could have deep significance to several of them. I highly recommend that this proposal be approved with full funding.

OVERALL RATING: ☒ EXCELLENT ☐ VERY GOOD ☐ GOOD ☐ FAIR ☐ POOR

NATIONAL SCIENCE FOUNDATION

PROPOSAL EVALUATION FORM

NSF Form 1 (4/84)
Supersedes All Previous Editions

PROPOSAL NO.	INSTITUTION	PLEASE RETURN BY
DCR-8613574	Stanford Univ	07/19/86

PRINCIPAL INVESTIGATOR	NSF PROGRAM
John M. Chowning	INTELLIGENT SYSTEMS PROGR

TITLE: Intelligent Analysis of Composite Acoustic Signals

Please evaluate this proposal using the criteria presented on the back of this review form. Continue on additional sheet(s) as necessary.

This proposal deals with the analysis and perception of acoustic signals by computers. Specifically, the investigators propose a general framework for the recognition of complex sounds from multiple sources. The project represents a continuation and extension of work performed at the CCRMA over the past five years.

On the positive side, I found the proposed research important and exciting. It attempts to develop a computational model for human sound perception, utilizing disciplines ranging from acoustics and music to computer science and signal processing. The proposed architecture enables investigators to represent knowledge and constraints from different domains in an independent manner, while providing a flexible mechanism for them to interact. The investigators are competent, and the facility more than adequate. Thus the project has the potential of advancing significantly our knowledge about human intelligence.

A number of issues, however, remain unresolved in my mind. The proposal represents a major departure from the investigators' previous effort in that they now stress the domain-independent nature of their research, moving away from music to encompass other acoustic signals. I believe such a change in emphasis could be a mistake, since domain-specific knowledge can provide the necessary leverage for sound perception. The motivation for such a shift in direction is not very clear. To me, this is not unlike proposing to build a speech recognition system that is language-independent. I am also dubious of their ability to meet their goal of dealing with polyphonic sound, since I saw no clear evidence that their research results can can successfully generalize from a single source to multiple sources. I believe the investigators are overly ambitious and optimistic; they have outlined a far too difficult problem to tackle in a two-year period. Finally, I am unable to tell whether the research emphasis is on the understanding of sound perception, or on the development of an expert system. Perhaps a more specific research plan would have helped.

OVERALL RATING: ☐ EXCELLENT ☑ VERY GOOD ☐ GOOD ☐ FAIR ☐ POOR