Image Cover Sheet

CLASSIFICATION	SYSTEM NUMBER 148988
UNCLASSIFIED	

TITLE

PROPOSALS FOR ENHANCING THE AUDITORY PRESENTATION OF SONAR INFORMATION

System Number:

Patron Number:

Requester:

Notes:

DSIS Use only:

Deliver to: FF

National Defence | Défense nationale

Canada

Defence and Civil INSTITUTE OF ENVIRONMENTAL MEDICINE

INSTITUT DE MEDECINE ENVIRONNEMENTALE pour la défense

1133 Sheppard Avenue West, PO Box 2000, North York, Ontario, Canada M3M 3B9
Tel. (416) 635-2000 Fax. (416) 635-2104

November 1994 DCIEM No.94-52

PROPOSALS FOR ENHANCING THE AUDITORY PRESENTATION OF SONAR INFORMATION

S. M. McFadden
R. Arrabito

Defence and Civil Institute of Environmental Medicine
1133 Sheppard Avenue West, P.O. Box 2000
North York, Ontario
Canada M3M 3B9

Table of Contents

ABSTRACT

Since the Second World War, visual displays have increasingly replaced auditory displays for the presentation of passive sonar sounds. Visual displays have proven more suitable for processing the low frequency, long duration, narrowband sounds picked up by passive sonar systems. However, the ability to detect and classify transients has been degraded as a result of this change. Moreover, the visual modality is becoming increasingly overloaded. As a result, there is interest in reviving the auditory display. However, an auditory display is not likely to be useful unless methods are developed for enhancing the the incoming data before presenting it aurally. Based on existing knowledge of human auditory capabilities and the passive sonar environment, possible enhancements include filtering, frequency shifting, and temporal compression to improve the detectability of low frequency, narrowband sounds, temporal compression and binaural presentation to improve the efficiency of processing multiple beams, and cueing to assist the recognition process. Most of these are feasible, but there is little direct research on their utility for improving the processing of passive sonar information aurally. A research programme is proposed for evaluating the different enhancements.

INTRODUCTION

Originally, the information received by hydrophones was presented to sonar operators via the auditory modality. However with the development of modern computers and visual displays, it was possible to process the incoming sounds and present them visually. The algorithms used to analyze the sounds proved ideal for processing narrowband low frequency information. Furthermore, operators required less training on the visual display than the auditory display, inaudible low frequency information could be displayed, and performance was not affected by high ambient noise in the environment. Consequently, the use of auditory displays for passive sonar declined (1). However, the human visual sense compares poorly to the auditory sense in interpreting certain information such as transients (2, 3). As ships become quieter, it is important that the operator be able to search for all sources of sound that may be emitted by a ship and thus the operator could benefit by simultaneously processing passive sonar data in both the auditory and visual modalities.

There are other reasons for including an auditory display component in the Human Machine Interface (HMI) of a passive sonar system. Increases in computing power have resulted in a requirement for the sonar operator to process larger quantities of information. Most of this information is presented visually, usually on some form of electronic display. This trend has resulted in the need to keep track of multiple pages of information and an overloading of the visual modality. This problem applies to most complex computer-based systems and has led to a renewed interest in developing auditory displays that could offload the visual modality. Thus, an aural display should be seen not only as a means of accessing specific types of information but also as a means of reducing operator workload and increasing efficiency.

In order to meet these requirements, careful consideration will have to be given to the form an auditory display should take. Most of the emphasis in the development of sonar displays has been on improving the presentation of information to the eye. Auditory displays still tend to present unprocessed data or data processed to optimize the visual display. Given such an emphasis, it is not surprising that performance on visual displays is superior to that found on auditory displays. To produce an effective and useful auditory display, one must develop methods for enhancing the incoming data in ways that are consistent with the auditory capability of the operator.

HUMAN AUDITORY CAPABILITY

Several reviews of human auditory capabilities as they apply to the processing of sonar sounds have been carried out (4-7). Based on these reviews, the ear is best suited to the detection and recognition of short duration, broadband, rapidly changing sounds such as speech. For example, detection and discrimination thresholds are lowest for frequencies in the speech range. Detection thresholds are lowest between about 500 and 5000 Hz; frequency differences of about 0.1% can be discriminated for frequencies up to 3KHz and intensity differences of 0.5 dB can be discriminated for frequencies up to 8KHz. Both these discrimination thresholds apply only under high signal to noise ratios. Threshold increases at higher frequencies and low signal to noise ratios. The ear is also sensitive to time and intensity delays between the two ears. Interaural intensity differences of 0.5 dB are perceptible between 0.2 and 10 KHz while time delays as small as a few microseconds are perceptible under the right conditions if most of the spectral content of the sound is below 1400 Hz. This capability is used for localizing sounds in space.

Good detection of broadband sound is possible because the ear is able to integrate energy from different parts of the spectrum and to separate out sounds arriving from different sources. On the other hand, any energy within a sixth to third octave band around a tone or narrow band noise will reduce the detectability of that sound. This limits the effectiveness of the ear for processing narrowband sounds. The integration period of the ear is less than a second. Thus, listening to a sound over a prolonged period does not lead to improved detectability.

The ear's frequency range, short integration period, and coarse filter bank means that it is poorly suited to the detection of the long duration, low frequency, narrowband sounds that populate the sonar environment. These sounds can be detected on visual displays because of the transforms used in processing information for that modality. Those same transform are not suitable for processing broadband transient sounds that the ear is most sensitive to. Thus it would seem that the auditory modality complements the visual modality in the processing of sonar information.

ENHANCEMENTS FOR AUDITORY DISPLAYS

Three types of enhancements can be considered for auditory displays. The first transforms the long duration, low frequency, narrowband sounds so that they are more easily processed by the ear. Possible transformations include filtering, frequency shifting, and temporal compression. The second type of enhancement involves more efficient processing of the incoming sounds. This might be accomplished through temporal compression or simultaneous presentation of multiple channels. The third type of enhancement is aimed at improving the operators ability to classify or categorize the incoming sounds.

Filtering

As stated above, the ability of the ear to detect a signal is reduced when it occurs in conjunction with other sounds that have a similar frequency spectrum or are lower in frequency. A straightforward method for improving detection under such conditions is to filter out the interfering frequencies by passing incoming sounds through an adaptive filter. Such a filter has been developed by Diagnostic/Retrieval Systems (6) and used successfully to improve aural detection on active sonar systems and the detection of torpedoes on passive systems. Its primary purpose is to improve the aural detection of long-duration tonals. Since similar information is available visually, the enhanced aural display would provide redundant information to the operator improving his confidence that a target is or is not present.

Frequency Shifting

Often, most of the energy in interesting signals occurs at frequencies below the frequency range of the ear. Long duration narrowband signals of this type can be picked up on visual displays, but transients signals are usually not seen. Shifting the incoming sounds upwards in frequency should make transients more detectable aurally. It is critical, however, that the whole frequency range be maintained. Otherwise the higher frequency components in the original sound may be lost, potentially reducing detection and discrimination.

Experienced operators would have to be trained to recognize the transformed sounds. This may be straightforward for sounds that do not have any frequency content above 200 Hz because the operator does not have templates for such sounds. Targets with some high frequency components may be more difficult because the operator would have to overcome existing stereotypes. Even naive operators will have certain stereotypes about engine sounds and marine sounds that may interfere with recognizing the transformed sounds. It would probably be best to allow the operator to listen to the sounds with and without frequency shifting and to provide examples of targets unshifted and shifted in frequency.

Temporal Compression

Most passive sonar targets have durations from a few seconds up to many minutes. Presenting such sounds for their entire duration will not make them more detectable to the ear which has an integration period of less than a second. Moreover, the ear tends to tune out sounds that do not change over time and to attend to any sudden changes in the acoustic environment. These factors suggest that detection of targets might be improved by compressing the incoming signals. Temporal compression would speed up slow changes in frequency and amplitude and potentially make signals more perceptible aurally. Temporal compression has several other advantages. It would allow the operator to listen to the same information repeatedly or to several channels or directions (beams) in the time period that he or she would normally listen to the uncompressed sound from a single channel. Thus, time compression would serve not only as a method for improving detectability, but also as a method for surveying different channels without the increase in detection threshold that occurs with visual summary displays. With modern processing techniques, it is possible to manipulate temporal compression and frequency shifting independently. Thus, compressing by factors of 10 to 100 would be possible without changing the frequency range of the sounds. The characteristics of such a display are discussed by McFadden and Taylor (5).

Three Dimensional Audio

Humans hear sounds in three dimensions; that is, sounds can be perceived as occurring above, below, in different directions, and at different distances from the listener. Using current knowledge of human binaural hearing, monaural sound can be transformed so that, when presented over headphones, the listener imagines that the sound is occurring at a specific location outside of the head. Unlike a dichotic presentation, a three dimensional presentation (3D) can convey intimacy or urgency when sounds approach the listener.

A 3D audio presentation of passive sonar signals may improve the operator's processing efficiency. Two or more channels can be presented simultaneously to the operator at different locations. This technique could have several advantages. As with temporal compression, 3D audio would allow multiple channels to be processed in a shorter time without the concomitant reduction in resolution that occurs with multiple channel visual displays. As well, the operator can compare the output of nearby channels which could enhance the perceptibility of a near threshold target that occurs in one channel and not the others. Alternatively one could present examples of known targets in one spatial location and output from one or more channels at the other spatial locations in the listener's head. The operator could then evaluate potential hypotheses about the target's classification.

Cueing

A common problem with auditory displays is the extensive training required before an operator can make effective use of them. Training is necessary for the operator to learn the many patterns and variations on patterns that can occur. It may be possible to reduce this training time or compensate for lack of training through the use of a dichotic presentation mode and a library of known target sounds. With a dichotic mode, the operator could listen to two different channels simultaneously (one in each ear) or listen to one channel in one ear and selected samples from a library of targets in the other ear. The first presentation mode turns the detection task (is there anything interesting out there) into a discrimination task (are these two channels the same or different). The thresholds for a two alternative forced choice detection (which involves a comparison amongst two alternatives) is lower than the threshold for a yes/no task (is there anything out there) (8). Similarly, the availability of a library of sounds compensates for the lack of an internalized library that possibly gives the experienced operator his or her advantage.

Usefulness of the Proposed Enhancement Techniques

To the best of our knowledge, very little research has been carried out evaluating the usefulness of any of the above techniques to the auditory processing of passive sonar information. Most of the transformations are technically feasible; however, it is not known if they would improve the perceptibility of passive sonar information. Diagnostic/Retrieval Systems claimed that they could get improvements of 17 to 25 dB with their audio enhancement system which incorporated filtering, frequency shifting, and frequency expansion (to improved detection of doppler shifts) (6). However, their system has been used primarily with active sonar. The gains that they achieved in detecting and classifying passive sonar targets were not specified in the reports available to us.

An unpublished study carried out at DCIEM indicated that temporal compression could enhance detection, but we have not investigated its suitability for rapid monitoring of multiple channels.

Three-dimensional reproduction has and is being considered for a wide range of applications. Some preliminary research has shown that binaural (3D) representation has several advantages over dichotic reproduction. The areas of research include speech intelligibility, spatial orientation, motor guidance, and teleconferencing. It must be emphasized that the usefulness of a 3D auditory display is greatly dependent on the suitability of the head transfer functions used to convert the sounds into 3D space, and the individual's ability to localize the sounds in 3D auditory space. To our knowledge, there are about five commercial products on the market. These are primarily tailored to deliver speech sounds and the spatial localization that would be achieved with passive sonar sounds remains to be determined.

With all of the proposed enhancements, it is important that the auditory display covers a wide frequency range in order to take advantage of the spectral integration capability of the ear. Also the display should cover a wide dynamic range in order to optimize frequency and intensity discrimination. This means using a reasonably high sampling rate and number of bits per sample. There are other reasons for providing as high a sampling rate as possible. Experience with digital audio suggests that under sampling can produce distortions in the sounds. Digital sound theory says that the sampling rate should be twice the highest frequency reproduced. However, such a sampling rate will be acceptable only if the sound is then passed through a filter with very sharp skirts. Otherwise there will be audible aliasing. Moreover, these distortions, unlike ambient noise due to other targets or sources, tend to distort the signal which could affect recognition. The ear can separate out sounds as a function of their source and disregard noise from uninteresting sources. However, the distortion due to undersampling is perceived as coming from the same source as the signal of interest. Most designers recommend a sampling rate of at least four times the highest frequency. A discussion of the errors associated with digital representations of sounds and the implications of these errors can be found in a report by Pennycook (9).

RECOMMENDATIONS

Independent of the utility of any of the transformations outlined above, it should be possible to develop a useful auditory display. The simplest form of display is one that would allow the operator to listen to two different channels simultaneously or one channel in one ear and a set of known archtypes in another ear. This display should be designed to emphasize the classification of transients. Thus, it would have a wide bandwidth to maximize frequency integration and a wide dynamic range to ensure good discrimination. The detection and classification of long duration sounds is handled quite adequately via the visual modality. Thus, it seems most sensible to concentrate initially on developing an auditory display that presents those types of sounds that the visual display is not suitable for. Moreover, research by Halpern (10) showed greater improvements in performance with a bimodal display when the information was complementary rather than redundant.

Having produced a display with which the operator can hear (detect and classify) known transients that are audible in the frequency range of the ear, the next step would be to provide tools that the operator could use to enhance the perceptibility of these types of sounds and similar sounds that are below the frequency range of the ear. This will probably be accomplished through some type of filtering and frequency shifting.

A third project would be to investigate the usefulness of an auditory display for rapid monitoring of different channels or beams. This display would present a time compressed output (with or without some frequency shifting) of each beam in succession or several beams simultaneously at different locations in the head. This type of display, although technologically feasible, has the most unknowns associated with it in term of ultimate value. If it works, it should certainly provide a more efficient method of monitoring the output of a towed array. However, we do not have any data on the level of time compression that would be useful or how many channels might be presented simultaneously.

FUTURE RESEARCH AND DEVELOPMENT

Development of the initial display suggested above is straightforward and could be initiated immediately if funding were available. However, there are some characteristics of that display that should be investigated either prior to or during the development phase. One such issue is what is a suitable sampling rate and signal to noise ratio. DCIEM has the capability to look at the impact of sampling rate and number of bits on the detection, discrimination, and classification of complex non-speech sounds. However, any results in our laboratory would have to be verified in the field.

A second issue is the impact of an auditory display on the operator's processing efficiency. Most studies on bimodal processing measure the detection or classification of single targets. With a bimodal display, the operator's monitoring load is potentially increased. Any improvements in performance through the use of a bimodal display may be negated by the increased processing load. There is a requirement to assess the advantage of bimodal displays under different levels of workload.

Similarly, providing the operator with archtypes should improve his or her ability to recognize targets. However, it takes time to access these archtypes and under high workload conditions the operator may not have or take the time to use such a tool effectively.

The development of an auditory display incorporating any or all of the enhancements discussed above is a long term and reasonably high risk development project. Some of the concepts could be evaluated in a research environment. In fact, DCIEM is currently investigating the effect of providing a library of sounds on classification performance, performance on a bimodal display under high workload conditions and the feasibility of a 3D auditory display. As well, based on the latest information available to us, Diagnostic/Retrieval Systems has an ongoing interest in developing their adaptive filter to improve auditory processing of passive sonar information. If funding was available, it would be useful to develop a prototype system incorporating the capabilities discussed above that could be used with real sonar data. It would probably be best for such a development to be jointly monitored by DREA and DCIEM.

CONCLUSION

Based on existing knowledge of the auditory and visual system and passive sonar systems, it would seem useful to provide an auditory display as part of the human machine interface of CANTASS. Such a display should complement the visual display by optimizing the presentation of transient broadband sounds. At the very least, the display should allow a dichotic presentation of sounds so that an operator can listen either to two channels simultaneously or to one channel in one ear and an example of a target in the other. Other enhancements that might be provided are time compression, frequency shifting, filtering, and 3D. Such enhancements have the potential of improving the perceptibility of targets and the sonar operator's information processing efficiency. However, their actual value remains to be determined. Evaluation of the proposed enhancements could best be accomplished through a joint effort of industry and CRAD.

REFERENCES

1. KOBUS, D. A. and L. J. LEWANDOWSKI. *Critical factors in sonar operation: A survey of experienced operators*. NHRC Report No. 91-19, Naval Health Research Center, San Diego, CA, 1993.
2. URICK, R. J. *Principles of underwater sound.* McGraw-Hill Book Co., New York, NY, 1975.
3. SMITH, M. J. Making the Most of Towed Arrays: The Human Factor. *Journal of Naval Science.* **9**: 145-156, 1983.
4. MCFADDEN, S. M. *Auditory Display of Passive Sonar Information.* DCIEM No. 84-R-23, Defence and Civil Institute of Environmental Medicine, Downsview, Ontario, 1984.
5. MCFADDEN, S. M. and M. M. TAYLOR. Human Limitations in Towed Array Sonar Recognition. *Proceedings of the 24th DRG Seminar on The Human as a Limiting Element in Military Systems.* **1**: 431-451, 1983.
6. *Passive enhanced audio study*. Technical report, Diagnostic/Retrieval Systems Inc., Oakland, NJ, 1991.
7. MULLIGAN, B. E., L. S. GOODMAN, D. K. MCBRIDE, T. M. MITCHELL, T. N. CROSBY, D. P. GLEISNER, K. D. STEWART, and L. HITCHCOCK. *Human Engineer's Guide to Auditory Displays, Vol. 1: Elements of Perception and Memory Affecting Auditory Displays*. NAVTRAEQUIPCEN 80-D-0011/0037-1, Eagle Technology Incorporated, Orlando FL, 1984.
8. MACMILLAN, N. A. and C. D. CREELMAN. *Detection theory: A user's guide.* Cambridge University Press, New York, NY, 1991.
9. PENNYCOOK, B. W. *Perception of Sounds Regenerated from Digital Representations: A Survey of Current Literature*. Image and Audio Systems, Kingston, Ontario, 1988.
10. HALPERN, J. *Bisensory Information Processing: Discrimination as a Function of Cue Redundancy and Task Difficulty*. TR-6, Computer Image Corporation and University of Denver, Department of Psychology, Denver, CO, 1971.

SECURITY CLASSIFICATION OF FORM
(Highest classification of Title, Abstract, Keywords)

DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when the overall document is classified)

1. ORIGINATOR (the name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g., Establishment sponsoring a contractor's report, or tasking agency, are entered in section 12.) Defence and Civil Institute of Environmental Medicine	2. DOCUMENT SECURITY CLASSIFICATION (overall security classification of the document including special warning terms if applicable.) Unclassified

3. DOCUMENT TITLE (the complete document title as indicated on the title page. Its classification should be indicated be the appropriate abbreviation (S,C,R or U) in parentheses after the title.)

Proposals for Enhancing the Auditory Presentation of Passive Sonar Information

4. DESCRIPTIVE NOTES (the category of the document, e.g., technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

Technical Report

5. AUTHOR(S) (Last name, first name, middle initial. If military, show rank, e.g., Burns, Maj. Frank E.)

Sharon McFadden and Robert Arrabito

6. DOCUMENT DATE (month and year of publication of document) November 1994	7.a. NO. OF PAGES (total containing information. Include Annexes, Appendices, etc.) 10	7.b. NO. OF REFS. (total cited in document) 10

8.a. PROJECT OR GRANT NO. (if appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant) NA	8.b. CONTRACT NO. (if appropriate, the applicable number under which the document was written) NA
9.a. ORIGINATOR'S DOCUMENT NUMBER (the official document number by with the document is identified by the originating activity. This number must be unique to this document.)	9.b. OTHER DOCUMENT NO.(S) (any other numbers which may be assigned this document either by the originator or by the sponsor.)

10. DOCUMENT AVAILABILITY (any limitation on further dissemination of the document, other than those imposed by security classification)

X	Unlimited distribution
	Distribution limited to defence departments and defence contractors; further distribution only as approved
	Distribution limited to defence departments and Canadian defence contractors; further distribution only as approved
	Distribution limited to government departments and agencies; further distribution only as approved
	Distribution limited to defence departments; further distribution only as approved
	Other

11. ANNOUNCEMENT AVAILABILITY (any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (10.) However, where further distribution (beyond the audience specified in 10) is possible, a wider announcement audience may be selected.)
As above

12. SPONSORING ACTIVITY (the name of the department project office or laboratory sponsoring the research and development. Include the address.)

Unclassified

SECURITY CLASSIFICATION OF FORM
(Highest classification of Title, Abstract, Keywords)

SECURITY CLASSIFICATION OF FORM
(Highest classification of Title, Abstract, Keywords)

13. ABSTRACT (a brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual).

Since the Second World War, visual displays have increasingly replaced auditory displays for the presentation of passive sonar sounds. Visual displays have proven more suitable for processing the low frequency, long duration, narrowband sounds picked up by passive sonar systems. However, the ability to detect and classify transients has been degraded as a result of this change. Moreover, the visual modality is becoming increasingly overloaded. As a result, there is interest in reviving the auditory display. However, an auditory display is not likely to be useful unless methods are developed for enhancing the the incoming data before presenting it aurally. Based on existing knowledge of human auditory capabilites and the passive sonar environment, possible enhancements include filtering, frequency shifting, and temporal compression to improve the detectability of low frequency, narrowband sounds, temporal compression and binaural presentation to improve the efficiency of processing multiple beams, and cueing to assist the recognition process. Most of these are feasible, but there is little direct research on their utility for improving the processing of passive sonar information aurally. A research programme is proposed for evaluating the different enhancements.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible, keywords should be selected from a published thesaurus, e.g. Thesaurus of Engineering and Scientific Terms (TEST) and the thesaurus identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

passive sonar, auditory displays, filtering, temporal compression, 3-dimensional sound

HFD 07/94

Unclassified

SECURITY CLASSIFICATION OF FORM
(Highest classification of Title, Abstract, Keywords)