Copy No. _____

Defence Research and Development Canada Recherche et développement pour la défense Canada

DEFENCE R&D DÉFENSE

Aural Classification for Active Sonar

Final Report for TIF Project 11CQ11 "Aural Discrimination of True Targets from Geological Clutter"

Paul C. Hines

Nancy Allen

Victor W. Young

Defence R&D Canada – Atlantic

Technical Report

DRDC Atlantic TR 2007-361

May 2008

Canada

This page intentionally left blank.

Aural Classification for Active Sonar

Final Report for TIF Project 11CQ11 "Aural Discrimination of True Targets from Geological Clutter"

Paul C. Hines
Nancy Allen
Victor W. Young

Defence R&D Canada – Atlantic

Technical Report
DRDC Atlantic TR 2007-361
May 2008

Principal Author

Original signed by Paul C. Hines

Paul C. Hines

PM/Rapid Environmental Acoustics

Approved by

Original signed by Neil Sponagle

Neil Sponagle

Head/Underwater Sensing

Approved for release by

Original signed by James L. Kennedy

James L. Kennedy

Chair/Document Review Panel

This Report produced for DRDC Atlantic project 11CQ11

Abstract

Military sonars must detect, localize, classify, and track submarine threats from distances safely outside their circle of attack. Active sonars operating at low frequencies are favoured for the long ranges they afford against quiet targets. However, in littoral environments, echoes from geological features (clutter) are frequently mistaken for targets of interest which results in degraded performance. Conventional signal processing techniques for dealing with this have met with only limited success *and* ignore a potentially valuable discrimination tool – the human auditory system. For it is generally accepted within the sonar community that operators can aurally discriminate target echoes from echoes from naturally occurring objects; however, the performance improvement has never been quantified. This report presents the results from a DRDC Technology Investment Fund (TIF) project whose aim was to measure the ability of the human operator to discriminate clutter from true echoes, to identify specific aural features that the operator employs, and to integrate these features into an automatic aural classifier. Using the area under the Receiver Operator Characteristic (ROC) curve, A_z, as a measure of performance, the results show that both the human operators and the automatic classifier ($A_z \geq 0.98$) score near to ideal performance ($A_z = 1$) for a broadband impulsive source. Performance degrades somewhat when acoustic bandwidth is reduced but is still very good ($A_z \approx 0.8$). Furthermore, results indicate that both the operators and the automatic classifier were performing the classification in essentially the same way. As well as reviewing the project results, the future direction of the research is highlighted.

Résumé

Les sonars militaires doivent détecter, localiser, classifier et poursuivre les menaces sous-marines à des distances de sécurité à l'extérieur de leur cercle d'attaque. Les sonars actifs à basse fréquence sont préférables en raison de leurs longues distances de fonctionnement contre les cibles silencieuses. Toutefois, dans les environnements littoraux, les échos provenant d'éléments géologiques (clutter) sont souvent confondus avec des cibles d'intérêt, ce qui occasionne une dégradation du rendement. Les techniques classiques de traitement des signaux permettant de faire face à ce problème n'ont connu qu'un succès limité et ne tiennent pas compte d'un outil de discrimination qui pourrait s'avérer précieux : le système auditif humain. Bien qu'il soit généralement reconnu, dans les milieux intéressés par le sonar, que les opérateurs peuvent discriminer à l'oreille les échos des cibles et les échos des objets naturels, l'amélioration du rendement n'a jamais été quantifiée. Le présent rapport expose les résultats d'un projet du Fonds d'investissement technologique (FIT) de RDDC, qui vise à mesurer l'aptitude de l'opérateur humain à discriminer le clutter des échos vrais, à déterminer les caractéristiques auditives particulières employées par l'opérateur et à intégrer ces caractéristiques en un classificateur sonore automatique. Lorsqu'on utilise la zone située sous la courbe caractéristique de fonctionnement du récepteur (ROC), Az, pour la mesure du rendement, les résultats indiquent que les opérateurs humains et le classificateur automatique (Az ≥ 0,98) donnent un rendement quasi-idéal (Az = 1) pour une source d'impulsions à large bande. Le rendement se dégrade quelque peu lorsque la largeur de bande acoustique est réduite, mais il demeure néanmoins très bon (Az ≈ 0,8). Par ailleurs, les résultats indiquent que les opérateurs et le classificateur automatique procèdent à peu près de la même façon pour exécuter la classification. En plus de l'examen des résultats du projet, l'orientation future de la recherche est mise en évidence.

This page intentionally left blank.

Executive summary

Aural Classification for Active Sonar: Final Report for TIF Project 11CQ11 "Aural Discrimination of True Targets from Geological Clutter"

Paul C. Hines; Nancy Allen; Victor W. Young; DRDC Atlantic TR 2007-361; Defence R&D Canada – Atlantic. May 2008.

Introduction or background: This report presents the results from a DRDC Technology Investment Fund (TIF) project whose aim was to measure the ability of the human operator to discriminate clutter from true echoes, to identify specific aural features that the operator employs, and to integrate these features into an automatic classifier. In addition to the target classification aspects of the project, DRDC collaborated with the Geological Survey of Canada (Atlantic) to identify high-clutter and low-clutter areas on the Scotian shelf to support future trials that will examine active sonar performance in the littorals. As well as reviewing the project results, the future direction of the research is highlighted.

Results: Using the area under the Receiver Operator Characteristic (ROC) curve, A_z, as a measure of performance, the results show that both the human operators and the automatic classifier score near ideal performance for a broadband impulsive source. Performance degrades somewhat when acoustic bandwidth is reduced but is still very good. Results indicate that the automatic classifier is performing the classification *in a similar way* to the operators.

Significance: Military sonars must detect, localize, classify, and track submarine threats from distances safely outside their circle of attack. Active sonars operating at low frequencies are favoured for the long ranges they afford against quiet targets. However, in littoral environments, operational sonars frequently mistake echoes from geological features (clutter) with targets of interest. This results in high false alarm rates and degradation in sonar performance. Conventional approaches □ using signal features based on the echo spectra or using signal features derived from physics-based models of specific target types □ have had only limited success; moreover, they ignore a potentially valuable tool for target-clutter discrimination – the human auditory system. That said, even if one verifies that aural discrimination is effective, discriminating targets from clutter is labour intensive and requires near-fulltime effort from the operator. Since future military platforms will have to support smaller complements, and near-future operations will have to accommodate additional mission-specific forces, automation of on-board systems is essential. The technique is well suited to autonomous systems since a much smaller bandwidth is needed to transmit a classification result than to transmit raw acoustic data.

Future plans: The success of the TIF has resulted in an Applied Research Program entitled □Automatic clutter discrimination using aural cues□which aims to integrate the automatic aural classifier algorithm with the reverberation-bathymetry mapping developed in project 11CQ on Rapid Environmental Assessment (REA) into a detection-classification system in DRDC's sonar test bed (Pleiades). This should substantially reduce the operator overload that results from unacceptably high false alarm rates currently present in navy sonars. The automatic aural classifier will also be tested using passive transients collected from torpedoes, submarines and marine mammals to examine its performance as an automatic classifier for passive sonar.

Sommaire

Aural Classification for Active Sonar: Final Report for TIF Project 11CQ11 "Aural Discrimination of True Targets from Geological Clutter"

Paul C. Hines; Nancy Allen; Victor W. Young; DRDC Atlantic TR 2007-361; R & D pour la défense Canada – Atlantique. May 2008.

Introduction ou contexte: Le présent rapport expose les résultats d'un projet du Fonds d'investissement technologique (FIT) de RDDC, qui vise à mesurer l'aptitude de l'opérateur humain à discriminer le clutter des échos vrais, à déterminer les caractéristiques auditives particulières employées par l'opérateur et à intégrer ces caractéristiques en un classificateur automatique. Outre les aspects du projet liés à la classification des cibles, RDDC a collaboré avec la Commission géologique du Canada (Atlantique) pour déterminer les zones de fort clutter et de faible clutter sur la plateforme néo-écossaise à l'appui de futurs essais qui examineront le rendement du sonar actif sur les littoraux. En plus de l'examen des résultats du projet, l'orientation future de la recherche est mise en évidence.

Résultats: Lorsqu'on utilise la zone située sous la courbe caractéristique de fonctionnement du récepteur (ROC), Az, pour la mesure du rendement, les résultats indiquent que les opérateurs humains et le classificateur automatique donnent un rendement quasi-idéal pour une source d'impulsions à large bande. Le rendement se dégrade quelque peu lorsque la largeur de bande acoustique est réduite, mais il demeure néanmoins très bon. Par ailleurs, les résultats indiquent que les opérateurs et le classificateur automatique procèdent à peu près de la même façon pour exécuter la classification.

Importance: Les sonars militaires doivent détecter, localiser, classifier et poursuivre les menaces sous-marines à des distances de sécurité à l'extérieur de leur cercle d'attaque. Les sonars actifs à basse fréquence sont préférables en raison de leurs longues distances de fonctionnement contre les cibles silencieuses. Toutefois, dans les environnements littoraux, les sonars opérationnels confondent souvent les échos provenant d'éléments géologiques (clutter) avec des cibles d'intérêt. Il en résulte des taux élevés de fausses alarmes et une dégradation du rendement du sonar. Les techniques classiques, qui font appel aux caractéristiques des signaux fondées sur les spectres d'écho ou qui utilisent des caractéristiques de signaux tirées de modèles physiques applicables à des types de cibles déterminés, n'ont eu qu'un succès limité; de plus, elles ne tiennent pas compte d'un outil qui pourrait s'avérer précieux pour la discrimination cible-clutter : le système auditif humain. Cela dit, même s'il est établi que la discrimination auditive est efficace, la discrimination des cibles et du clutter demeure fastidieuse et nécessite les efforts de l'opérateur presque à plein temps. Comme les futures plateformes militaires devront être dotées d'effectifs réduits et que les opérations devront dans un proche avenir répondre aux besoins de forces supplémentaires pour des missions déterminées, l'automatisation des systèmes de bord est essentielle. Cette technique convient bien aux systèmes autonomes, car la transmission d'un résultat de classification exige une largeur de bande beaucoup plus restreinte que la transmission de données acoustiques brutes.

Perspectives: Le succès du FIT s'est traduit par un programme de recherche appliquée appelé « Discrimination automatique du clutter au moyen d'indices sonores », qui vise à intégrer

l'algorithme du classificateur sonore automatique au système cartographique de bathymétrie à réverbération mis au point dans le cadre du projet d'analyse environnementale rapide 11CQ, de façon à constituer un système de détection-classification pour le banc d'essai sonar de RDDC (Pléiades). On devrait ainsi constater une réduction substantielle de la surcharge de l'opérateur, résultant des taux de fausses alarmes beaucoup trop élevés qui caractérisent actuellement les sonars de la Marine. Le classificateur sonore automatique sera aussi soumis à des essais, au moyen de signaux transitoires passifs provenant de torpilles, de sous-marins et de mammifères marins, ce qui permettra d'étudier son rendement comme classificateur automatique de sonar passif.

This page intentionally left blank.

Table of contents

List of figures

Acknowledgements

The authors wish to acknowledge the other members of the TIF project both for technical contributions and insightful discussions throughout the project: Mr. Robert Arrabito of DRDC Toronto; Dr. John C. Osler, Mr. Jeff Scrutton, Mr. Doug Caldwell, Dr. Sean Pecknold, Mr. Paul Shouldice, Mr. Dan Graham, and Dr. Ron Kessel of DRDC Atlantic; Dr. Robert Courtney and Dr. Dave Mosher of the Geological Survey of Canada, Dr. Jim Pitton and Dr. Scott Philips of the Applied Physics Laboratory-University of Washington, and Dr. Doug Abraham of the Office of Naval Research, Washington DC. The authors also wish to acknowledge the DND personnel from ADAC(A), CFMWC, CFNOS, MOG5, PSU HALIFAX, TRINITY, and DRDC Atlantic who volunteered for the listening tests, and to their Commanding Officers for supporting their participation.

This page intentionally left blank.

1 Introduction

Military sonars must detect, localize, and track submarine threats from distances safely outside their circle of attack. Active sonars operating at low frequencies are favoured for the long ranges they afford against quiet targets. One approach to allow low frequency active sonars to detect and localize potential targets is to compare the sonar echo returns with the bathymetry data for the operational area. To do this, one first normalizes the returns by the range dependence of the received echo using a sonar propagation model. When the normalized echo is overlaid on a bathymetric plot for the area, potential targets show up as hot spots – levels that are substantially higher than the expected reverberation. This methodology has been successfully demonstrated in a number of trials[1]. However, the technique degrades significantly in coastal environs, partly because of the highly variable bathymetry and partly because sonars often measure strong coherent returns from the seabed at locations where bathymetric maps indicate benign bottoms. (See Figure 1.) This frequently occurs in littoral waters where, for example, ancient river channels and iceberg scours have been in-filled with layers of soft sediment. The high-frequency systems employed to measure bathymetry are unable to penetrate the seabed, so these complex sub-bottom features go uncharted. However, operational sonar signals can propagate into the seabed and scatter coherently from these features. This results in a substantial increase in false alarm rates and a corresponding reduction in performance. These unwanted coherent returns are referred to as geological clutter or simply clutter. Although clutter and true targets are indistinguishable using typical energy detectors it is generally accepted within the sonar community that they are aurally dissimilar. In fact, along with the wealth of anecdotal evidence, there is mounting experimental evidence [2] to suggest that echoes from man-made metallic objects sound very different than similar echoes from naturally occurring objects when listened to by a human operator. In the vernacular of acoustics, this is referred to as auralization. (No doubt, there are those in the sonar community who believe that the ability to aurally discriminate targets is a long established fact. However, scientifically quantifying how well operators can perform this task – generating a Receiver-Operating Characteristic or ROC curve, for example – is only now being done.)

Even if one verifies that aural discrimination is effective, discriminating targets from clutter is labour intensive and requires near-fulltime effort from the operator. Since, future military platforms will have to support smaller complements, and near-future operations will have to accommodate additional mission-specific forces, automation of on-board systems is essential. There have been attempts to develop automatic classifiers for active sonar using signal features based on the echo spectra, and by using signal features derived from physics-based models of specific target types. These conventional approaches have met with some success but they ignore a potentially valuable tool for target-clutter discrimination: the human auditory system. This report describes the results of a Defence R&D Canada – Atlantic (DRDC Atlantic) Technology Investment Fund (TIF) research project [3] whose dual aim was to quantify the aural performance of the sonar operator in a controlled experiment, and to develop an automatic classifier based upon specific aural cues or perceptual signal features employed in the human auditory system.

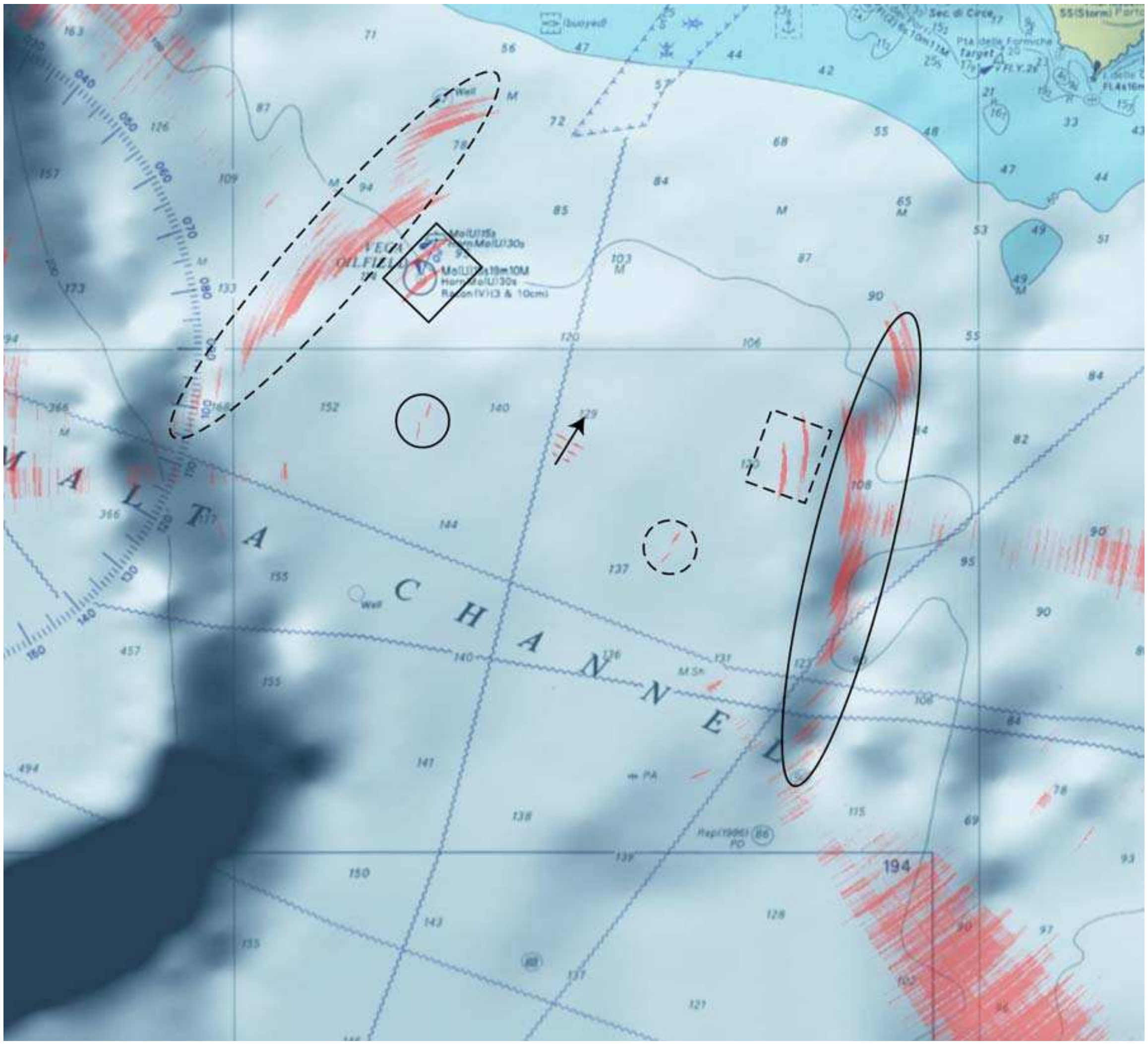

Figure 1: Reverberation-bathymetry composite plot used to isolate possible targets. The source and receiver are located at the center of the plot and the arrow indicates the direction of the towed array. Darker reds indicate higher levels of reverberation. Because of the left-right ambiguity of the array, the image is reflected on either side of the arrow. Solid lines enclose the real returns and dashed lines enclose virtual returns that result from the ambiguity. The oval encloses an area of high reverberation due to a rocky outcropping. The small circle encloses echoes from the seabed where the bathymetry is benign. The small rectangle encloses echoes from 2 surrogate targets to be discussed later in the text.

2 The Sonar Test Data

The first major hurdle to overcome in the project was to obtain a controlled data set. The importance of this cannot be overstated. The validity □and therefore the usefulness □of the results rests on a foundation built from the data. For example: one must ensure that all echoes used in the project have been correctly assigned to the target or clutter class; one must attempt to control variables such as SNR (signal-to-noise ratio) to ensure that operators aren□t classifying echoes based on SNR. These restrictions meant that operational data that is sometimes used for training would not be adequate. Instead, a specific experiment was performed during May 2004 to collect the data. The experiment was part of a joint trial undertaken by DRDC Atlantic, the Applied Research Laboratory of the Pennsylvania State University and the U.S. Naval Research Laboratory, and the NATO Undersea Research Center (NURC). The trial was conducted on the Malta Plateau, a region of the Mediterranean between the islands of Sicily, to the north, and Malta, to the south (See Figure 2). This region was selected as the experimental site because it is well surveyed, it is rich in clutter objects, and the water depths (typically 80-120 m) are consistent with CF navy littoral operations. Most importantly, the site provided 2 surrogate targets in the form of the Campo Vega oil platform and the tanker ship that attends it. These targets provided the necessary aural qualities but none of the logistical constraints inherent in using a real target.

It is intuitively apparent that aural classification requires a significant amount of acoustic bandwidth to be successful. That is to say, the echo of a 1 kHz tone may exhibit some time and frequency spreading as it reflects off an object but it will still sound essentially the same regardless of the reflecting object. Rather, the aural distinction occurs because different objects behave as complex filters, each possessing its own characteristic spectral response, amplifying and attenuating different frequencies within the spectrum of the ensonifying pulse. Because the minimum bandwidth required for aural discrimination was yet to be determined, an impulsive source was selected to ensure sufficient bandwidth. In all, nine signal underwater sound (SUS) charges (90 m detonation depth) were deployed during the sea trial to serve as broadband active sonar sources. SUS charges possess significant energy across their bandwidth and their waveforms and spectra are well characterized in the literature [3]. The nine charges were deployed one at a time from the stern of the Canadian Forces Auxiliary Vessel (CFAV) Quest. The charges were deployed in a relatively tight cluster (*Q* in Figure 2) at a range of approximately 13.5 km, south-southeast of Campo Vega. The active sonar receiver for the experiment was a towed array consisting of 96 omni-directional elements, with 0.5 m spacing between elements. The receiver array was towed behind CFAV Quest, the same ship from which the SUS charges were deployed, and so the experimental geometry was approximately monostatic. The receiver array was towed at a speed of roughly 6 knots at a depth of about 40 m and sampled at a rate of 4096 Hz.

Conventional time delay and sum beamforming was applied to the towed array data to form 81 horizontal beams, equally spaced in cosine. Because the towed array is composed of omni-directional elements there is an inherent left-right ambiguity in the beampattern. During the experiment, the ambiguity was overcome by using three different array orientations: three SUS drops each, on array headings of 266°, 146°, and 26°.

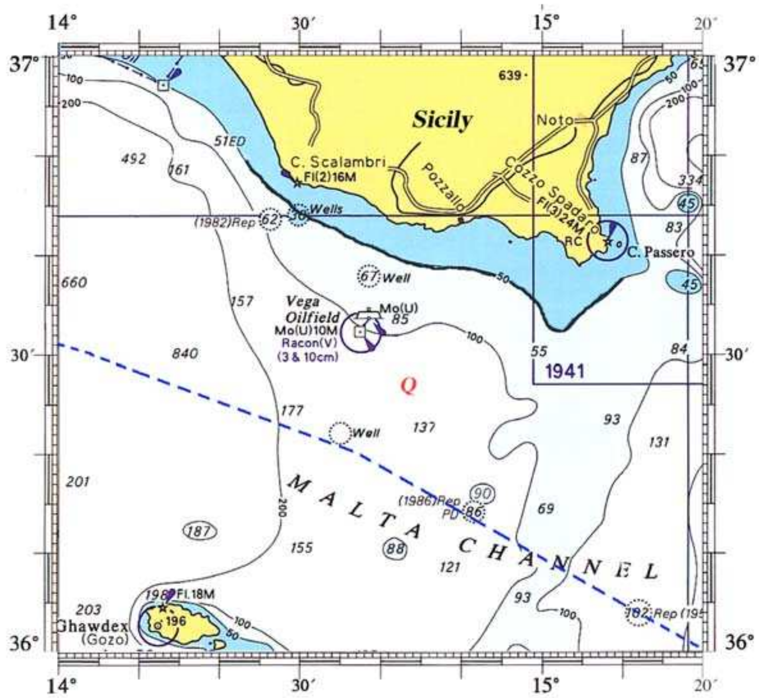

Figure 2: Experimental site for collecting auralization data. The red Q marks the centre point for the measurements.

Although this report focuses on the classification of sonar returns, considerable effort went into extracting suitable echoes (clutter and target) from the reverberation; that is, the detection process. The details are omitted here but the interested reader is referred to [5]. Briefly, the reverberation was normalized to account for propagation loss, and a threshold energy detector was applied. Each echo was associated with a location on the seafloor based on its time of arrival and beam angle. To remove spurious peaks from the energy detector, the echo must show up on at least two of the three array orientations to be considered a valid detection. Finally, the target and clutter echoes were extracted from the time series and stored for playback for the sonar operator and the automatic classifier. In all, the experiment produced 100 useable clutter echoes from 26 distinct objects, and 98 target echoes from the 2 surrogate targets [6].

All echoes were stored as individual .wav files in two forms: full bandwidth .wav files (100 - 2000 Hz), and reduced bandwidth .wav files (500 Hz – 2000 kHz). The motivation for using the reduced bandwidth is two-fold: First, the minimum bandwidth necessary for aural classification is of substantial interest and this provided some initial insight. Secondly, the goal is to eventually develop an aural classifier that employs a coherent projector (due to the environmental and logistical issues associated with explosive sources). Present technology limits coherent sources from operating much below 500 Hz so this provides an opportunity to see if using a coherent source would be feasible.

3 Quantifying Aural Performance of a Sonar Operator

Thirteen Canadian Forces sonar operators and 2 DRDC Atlantic civilians with sonar experience, participated in the listening tests which took place in a quiet room at DRDC Atlantic. The main objective of the test was to evaluate if humans could aurally discriminate between target echoes and environmental clutter, and if so, how well. Figure 3 contains a photograph of Maj. Robert Schwartz demonstrating one of the rating exercises. The experiment consisted of a training and a testing component. Following a brief familiarization exercise, a training phase occurred during which the subject listened to a subset of the echoes and rated them as clutter or target. After each decision, the correct answer was displayed to train the participant. In the testing phase, the subject was given an echo and asked to rank it along a 7 position scale ranging from “definitely or almost definitely clutter” to “definitely or almost definitely a target”. As well as rating the echoes in terms of this scale, the participants completed a qualitative questionnaire to provide insight into their decision making process. Further details on the listening test are contained in [7]. The quantitative results from the study can be used for evaluating how well operators perform the classification task. Both qualitative and quantitative results may point to new ways of tackling classification. The listening experiment was performed using the full-bandwidth data first. Approximately 6 weeks later the experiment was repeated using the reduced-bandwidth data.

Figure 3: Demonstration of the auralization set-up for the listening test.

The main measure of performance used to evaluate the listening tests was the Receiver-Operating Characteristic (ROC) curve. The ROC curve plots the probability of detection vs. probability of false alarm; that is to say, the probability of correctly identifying a target vs. the probability of labelling a clutter echo as a target [8]. One of the most useful metrics one can extract from the ROC curve is the area under the curve, A_z. The greater A_z, the better the classifier with a value of 1 indicating ideal performance. Three ROC curves are shown in Figure 4. The thin diagonal line represents chance performance; that is, a classifier that randomly classified echoes as either target or clutter which has an $A_z = 0.5$. The ROC curve corresponding to the average test subject for the full bandwidth data is given by the dashed curve and has an $A_z = 0.94$. The solid curve is the ROC for the top performer in the full bandwidth listening test and corresponds to an $A_z = 0.998$. Clearly, these values are indicative of a very successful classifier. For the reduced bandwidth test, the top performer obtained a value of $A_z = 0.91$ (dot-dash curve) and the average performer obtained $A_z = 0.77$ (dotted curve). As expected, performance has degraded with decreased bandwidth but the results are still impressive. Complete details on the scientific results of the listening test are contained in [9].

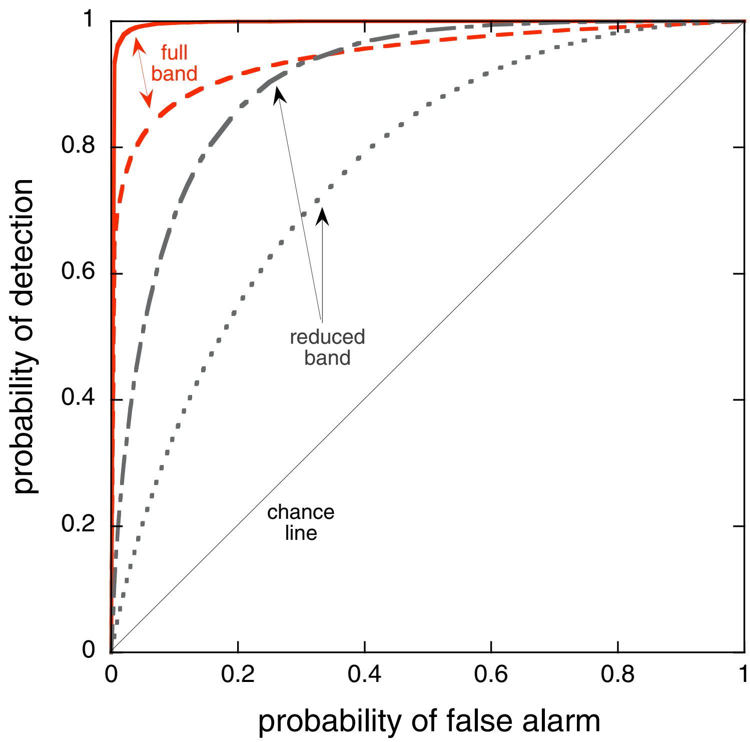

Figure 4: Sample ROC curves for the full-band listening tests for the top performer (solid line) and the average performer (dashed line). For the reduced bandwidth, the top performer is given by the dot-dash curve and the average performer is given by the dotted curve. The chance performance (thin diagonal line) is plotted for reference.

4 An Automatic Classifier Modeled on the Human Auditory System

To develop an automatic classifier based on the human auditory system one must shift away from the sonar paradigm modelled on conventional signal processing, and think in terms of perceptual features and the processes associated with human perception. This can best be understood by a simple example: in conventional signal processing one might employ sound intensity, that is, the total energy integrated across the entire frequency band of the signal; this is not the same as the perceived loudness of the same signal since the human ear applies a bandpass filter (maximum response at about 1 kHz) before the brain even processes it into something we register as sound. The question to address then is "*How can one extract perceptual features from the sonar echoes for use in an automatic classifier*?" This is examined next.

It is possible to draw an analogy between explosive-source active sonar echoes and the sounds produced by percussive musical instruments. This may not seem particularly intuitive; however, we note that both types of sounds are generated by exciting a physical object with an impulse. Motivated by this analogy, we examined perceptual signal features that have previously been identified as important to the perception of musical timbre. Timbre refers to those attributes of a sound that allow a listener to discriminate between two sounds that have the same pitch and loudness. Simply put, it is timbre that allows us to discriminate between a piano and a guitar both playing middle C (the pitch) even when they are played at equal volumes (the loudness).

The automatic aural classifier (AAC) can be separated into 3 stages. In the first stage, the echoes are signal conditioned to approximate the response of the human ear. This processing is adapted from established mathematical models of the ear's response. Secondly, the values for a set of perceptual features are computed from the data. Two major types of perceptual signal features are considered: time-frequency features and purely spectral features. There are no purely temporal perceptual signal features in the human auditory system, because the cochlea breaks incoming signals down into sub-bands, so that unfiltered time-domain signals are not directly available to the brain. An example of an important spectral feature is the loudness centroid – the frequency of the "balancing-point" of the loudness spectrum. Note that this frequency is different from what one would measure using conventional signal processing because the ear filters the signal prior to passing it on to the brain for final processing and classification. An example of an important time-frequency feature is the sub-band correlation – that is, the correlation between various sub-bands of the echo. These sub-bands approximate the frequency discrimination of the basilar membrane inside the cochlea. It is this sub-band synchronicity that captures the importance of bandwidth in the auralization process. The ear–brain processor automatically correlates the echo across these sub-bands which plays a critical role in our ability to classify sounds. Finally, the AAC is trained on a subset of echoes much as the sonar operators were, as outlined in the previous section.

There are many legitimate ways that the full data set could be split into training and testing subsets and no one approach is inherently better than any other. The only constraint on the training-testing split is that the same echo may not be included in both subsets. In the typical operational setting, an automatic classifier is trained on an archival database of target and clutter echoes and is then tested on echoes that result from different target and clutter objects. In order to approximate the operational requirement, this paper considers a training subset composed of echoes from the tanker and half of the clutter objects, and a testing subset composed of echoes from the Campo Vega oil platform and the remaining clutter objects. This training-testing split is not quite as challenging as would be expected operationally – where the objects in the training and testing subsets may be situated in totally different environments; however, it does force classification to be carried out based on inherent similarities between the echoes from the man-made objects and inherent similarities between the echoes from the different clutter objects, rather than simply recognizing different echoes from the same object.

Two metrics were used to examine the performance of the automatic aural classifier: the ROC curve described earlier, and principle component analysis (PCA). To facilitate comparison with the human operators, the AAC output was treated as if it were the result of an additional participant in the listening test to compute the statistics. The ROC curve corresponding to the AAC for the full bandwidth is given by the solid line in Figure 5 and has an $A_z = 0.98$. The ROC curve corresponding to the *reduced* bandwidth is given by the dashed line and has an $A_z = 0.86$. These values are indicative of a very successful classifier. The lower limit of each shaded area corresponds to the average performer and the upper limit of each shaded area corresponds to the top performer as presented in Figure 4. It is worth noting that for both bandwidths used, the automatic classifier is performing well above average and very nearly as well as the top performer. This is particularly significant since the automatic classifier would not be subject to stress, fatigue, or weather-related performance degradation in an operational situation.

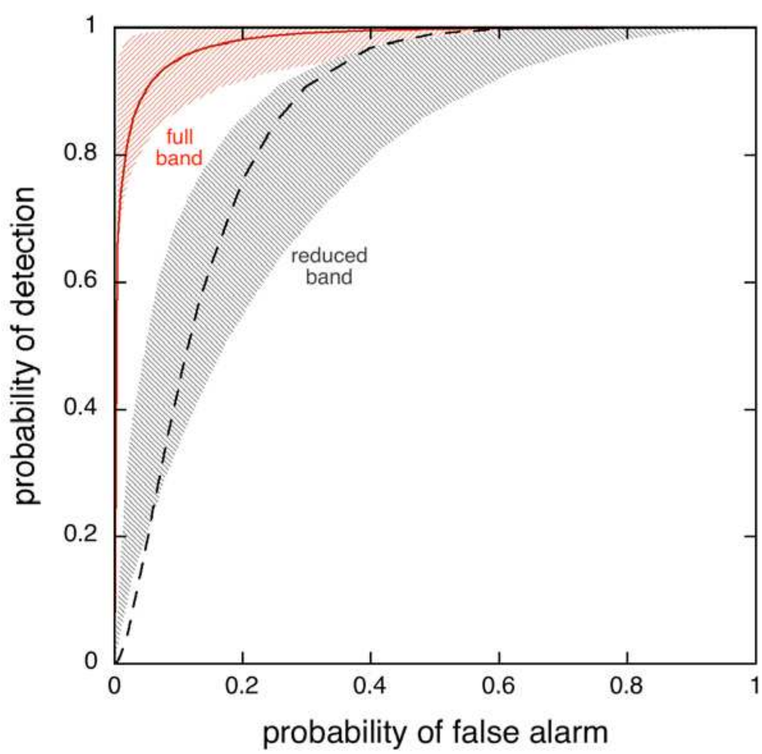

Figure 5: Comparison of ROC curves for the automatic aural classifier with the listening test results. The shaded regions mark the limits of the average and the top performer.

Because the mathematical details of PCA are somewhat cumbersome, only the results will be presented here. The interested reader is referred to [5] for additional details. Figure 6 contains a plot of the output of the Automatic Aural Classifier. Note that the features that make up the horizontal and vertical axes do not refer to specific aural features (e.g. loudness or duration) but rather, are a mathematical combination of the most important aural features used in the classification process.

Essentially, the further the clutter points (blue circles) separate from the target points (red squares), the better the classifier performs. If a red square lies on the red background, the classifier has correctly identified a target. Conversely, if a red square lies on the blue background, the classifier has a missed detection. Similarly, if a blue circle lies on the blue background, the classifier has correctly identified a clutter echo but if a blue circle lies on the red background it has a false alarm. The solid and dashed curves in the figure represent different decision surfaces and represent the ratio, R, of the probability of the echo being clutter to the probability of it being a target. The choice of where to set the decision surface is determined by the cost associated with an error. For example, if the cost associated with a missed detection is the same as the cost associated with a false alarm then one would set R = 1. If on the other hand, the cost of a missed detection was severe, classifying a hostile submarine as clutter, say, then one might set R = 10, and reduce the risk of a missed detection, at the expense of allowing additional false alarms [10]; that is to say, it would have to be 10 times more probable that an echo is from clutter before being dismissed as clutter. For R = 10, one obtains a 90% probability of detection with only an 8% probability of false alarm. Once again, this is indicative of a very successful classifier.

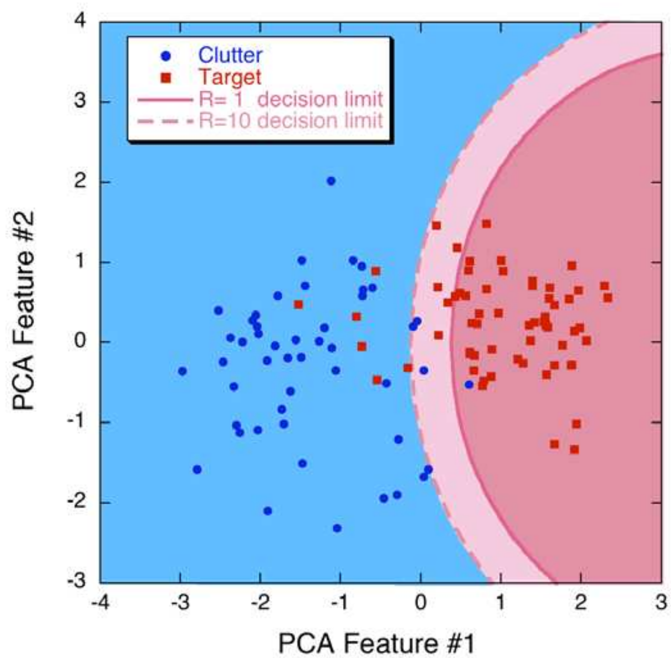

Figure 6: Output from the Automatic Aural Classifier (AAC). The choice of R is determined by the cost associated with an error (see text).

5 Quantifying the Human-Machine Link

In the previous sections scientific evidence was presented to show that human operators can aurally discriminate echoes from clutter from those from man-made metallic objects. We have also demonstrated that an automatic aural classifier based on musical timbre performs the task as well or better than the operators. The question remains, however; "Does the automatic aural classifier actually perform the task *in the same way* as the human operator?" I.e., has the model captured the features that humans use to classify the echoes? This is more than just of academic interest since, if the classifier is still not accurately capturing the features that humans use, additional performance enhancements may still be possible.

If the human operators and AAC are, in fact, performing the target-clutter classification in the same way, then the echoes that the human operators found easiest to correctly classify should be the same ones that the AAC found easiest to classify. Likewise, the echoes that the human operators found hardest to classify should be the same ones that the AAC found hardest to classify. Therefore, one way to quantify the human-machine link is to correlate the human operators performance with that of the AAC in terms of the classification difficulty of the echoes.

The correlation was performed using a simple ordinal ranking scale to quantify the human-machine link; that is, one in which the data points are ordered from smallest to largest, but in which the actual difference between individual points is disregarded. This method was chosen because the rating scale used in the listening experiments allowed subjects to order the echoes from most target-like to most clutter-like, but could not quantify the "separation" in perceptual space between categories. For example, it is possible to say that an echo with a rating of 2 received a smaller score than an echo with a rating of 3, but it is not possible to say that the difference between echoes receiving scores of 2 and 3 is the same as the difference between echoes receiving scores of 3 and 4. The Spearman rank correlation, r_S, calculates correlation on the *ranks* rather than the actual *values* and was used to quantify the human-machine link. To calculate r_S, target and clutter echoes were separately ranked [11] from "easiest to classify" to "hardest to classify" for both the human subjects and the AAC. In the case of the human operators, the echoes that were "easiest to classify" were those clutter echoes that were rated as "definitely or almost definitely clutter", and those target echoes that were rated as "definitely or almost definitely a target". Likewise, the echoes that were "hardest to classify" were those clutter echoes that were rated as "definitely or almost definitely a target", and those target echoes that were rated as "definitely or almost definitely clutter" (Refer to Figure 7.). In the case of the AAC the echoes that were "easiest to classify" were those clutter echoes that were located far from the PCA-space decision surface but on the clutter-side, and those target echoes that were located far from the decision surface but on the target-side and so on.

Based on a simple student's t-test, a rank correlation of $r_S > 0.22$ is required in order to claim statistical significance at a 95% confidence level for the number of degrees of freedom associated with the data set. In other words, to conclude that there is a relationship between the way that the human auditory system and the AAC perform target-clutter classification, it is necessary to observe a value of $r_S > 0.22$. When considered in isolation, eight classification features were observed to yield $r_S > 0.22$. Loudness centroid described earlier, for example, had $r_S = 0.45 \pm 0.11$. Thus the evidence indicates that the AAC is classifying in much the same way as the human operator. Even larger correlations could be achieved by considering multi-feature subsets but it was outside the scope of the project to investigate this further.

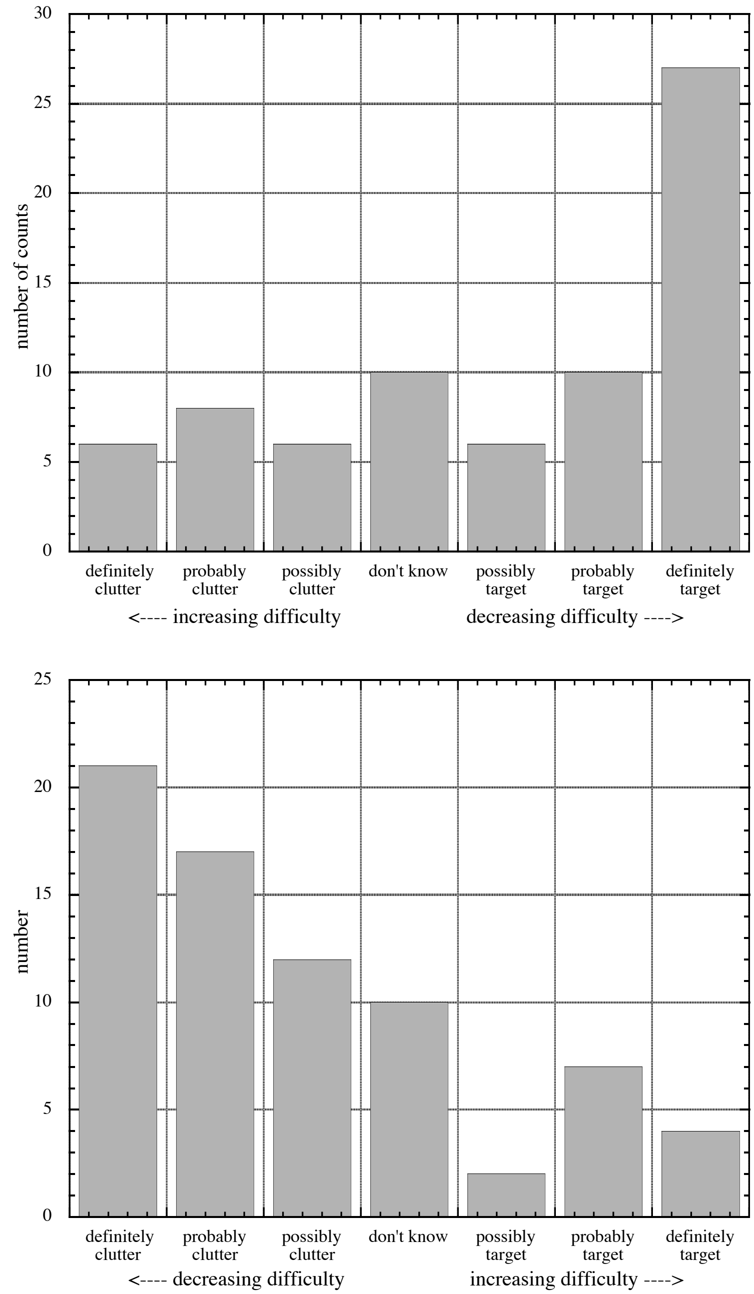

Figure 7: Histograms for the 7 classification categories ranging from "definitely clutter" to "definitely target" for the target echoes (upper figure) and clutter (lower figure) for subject S07 using the full-band data. Those identified as "definite" were ranked easiest to classify if they were correctly identified and hardest to classify if they were incorrectly identified, and so on.

6 Summary and Future Direction

The mandate of the Technology Investment Fund (TIF) is to investigate high risk research projects that have potentially high payoff. In that regard, the goal of the Aural Classification of Active Sonar TIF was to develop an automatic classifier for active sonar modelled after the human auditory system. This required sonar operator performance at the task to be quantified, key aural features used in the classification task to be identified, and an ear model to be implemented in software. At project completion, a prototype automatic classifier modeled after the human ear has been successfully demonstrated at DRDC Atlantic. The initial version of the system employed a broadband explosive source to ensure that sufficient bandwidth and source level was available to enable classification.

The success of the project has led to an Applied Research Project (ARP) whose goal is to incorporate the automatic aural classifier into the Canadian Navy's sonar suite using DRDC Atlantic's Pleiades sonar testbed. One of the first goals of the ARP is to demonstrate that similar performance can be obtained with a broadband coherent source because of the logistical and environmental difficulties associated with explosive sources. An experiment using a coherent source was conducted during the preparation of this report and initial results look very promising and will be the subject of future reports. The technique need not be limited to active sonar. In fact, active sonar can be considered the hard case since the theoretical foundation for this algorithm was actually based on passive musical acoustics; as a result, the ARP will examine whether the classifier can be modified to discriminate passive transients from submarines and torpedoes. This would allow covertness to be maintained and permit faster integration into the Canadian Navy. There is also a high public sensitivity to the impact of active sonars on marine mammals and indications are that this sensitivity will increase. This continues to have a strong negative impact on force generation since navy exercises at sea are a crucial component of training. An automatic discriminator that identifies marine mammal species would allow the CF to conduct operations, while fulfilling their mandate of environmental stewardship. Preliminary analysis of marine mammal vocalizations suggest that the classifier can be modified to discriminate them and the ARP will examine this in further detail.

Although outside the research scope of the current ARP, two other potential applications of the aural classifier should be noted. First, it is entirely reasonable to suggest that echoes from mine countermeasure (MCM) sonars will exhibit the same class specific nature as submarine targets, in spite of the higher frequencies encountered. That is, echoes from mines should be aurally distinct from echoes from boulders. There is evidence in the scientific literature to support this in which humans have aurally classified objects using high-frequency (>100 kHz) dolphin clicks by slowing the echo playback, to shift it to the audible frequency band [12]. Therefore, one could potentially employ the automatic aural classifier in an MCM role. Finally, one should be able to extend application to the land domain. For instance, one could envision a scenario in which ground troops deploy wireless audio transceivers for perimeter defence or area clearance and train the classifier to provide an automatic alert for specific classes of sounds such as voice, footsteps, or vehicle engines.

References

[1] Dale D. Ellis, John R. Preston, Reginald Hollett and Jurgen Sellschopp, "Analysis of towed array reverberation data from 160 to 4000 Hz during rapid response 97," SACLANTCEN, SR-280, 2000.

[2] N. Allen, P. C. Hines, V. W. Young and D. A. Caldwell, "Study on the human ability to aurally discriminate between target echoes and environmental clutter in recordings of incoherent broadband sonar," J. Acoust. Soc. Am. 119, 3395 (Abstract) (2006).

[3] A separate aspect of the project examined how well one could predict the impact of clutter on a sonar system. The Geological Survey of Canada were contracted to examine their data archives and construct maps of the Scotian shelf detailing potentially high-clutter and low-clutter regions, based on the seabed's composition and bathymetry. The maps are contained in a report authored by John Zevenhuizen, "Compilation and interpretation of Geophysical and geological data on the Scotian Shelf in support of summer 2004 acoustic geoclutter field program," Orca Marine Geological Consultants Ltd., 2004. Experimental verification of these predictions has not yet been performed.

[4] N. Ross Chapman, "Source levels of shallow explosive charges," J. Acoust. Soc. Am. 84, 697-702 (1988).

[5] Young, Victor W. and Hines, Paul C., "Perception-based automatic classification of impulsive-source active sonar echoes," J. Acoust. Soc. Am. 122 (3) pp. 1502-1517, 2007.

[6] Using 9 SUS on two targets only yields 18 returns. The additional target echoes were generated by using off-beam echoes. Clearly additional target echoes generated in this way are not completely independent since they result from a single physical process; however, differences in the time series obtained from steering in different beam directions, coupled with differences in ambient noise on different beams resulted in aurally distinct echoes.

[7] Nancy Allen and Paul C. Hines, "Protocols for Two Human Performance Tests on Aural Discrimination of Sonar Echoes," DRDC Atlantic Technical Memorandum, TM 2006-302, February 2006, 74 pages.

[8] Use of the ROC curve isn't limited to target detection, or even to military applications. The curve is used in a variety of disciplines where classification statistics are studied. For example, in medical diagnosis it is used to study the success rate of detecting cancer, in which case a false alarm might correspond to a benign growth being misdiagnosed as malignant.

[9] Nancy Allen, "Receiver-Operating Characteristic (ROC) Analysis Applied to Listening-test Data," DRDC Atlantic Technical Memorandum, TM 2007-353, August 2008, 73 pages.

[10] It is important to realize that one cannot indiscriminately increase R to avoid missing detections since an excessive false alarm rate will also degrade performance.

[11] If a single correlation ranking was performed on a combined set of target and clutter echoes, one obtains an erroneously high correlation since the result is dominated by the ability of listeners and machine to perform the discrimination, rather than by how the discrimination is performed.

[12] Whitlow W.L. Au and Douglas W. Martin, □nsights into Dolphin Sonar Discrimination Capabilities from Human Listening Experiments□, J. Acoust. Soc. Am. 86, 1662-1670 (1989).

List of symbols/abbreviations/acronyms/initialisms

AAC	Automatic Aural Classifier
MCM	Mine Counter Measures
NURC	NATO Undersea Research Centre
Pleiades	A DRDC test bed for advanced sonar integration and display
ROC	Receiver-Operating Characteristic
SNR	Signal-to-noise Ratio
SUS	Signal Underwater Sound

This page intentionally left blank.

Distribution list

Document No.: DRDC Atlantic TR 2007-361

LIST PART 1: Internal Distribution by Centre:

8	Paul C. Hines
5	DRDC Atlantic Library
1	H/US
1	H/TD
1	Nancy Allen
1	Jeff Scrutton
1	GL/MEA
1	Vince Myers
1	Garry Heard
1	Senior Military Officer
21	TOTAL LIST PART 1

LIST PART 2: External Distribution by DRDKIM

Department of National Defence

1 DRDKIM

Following 6 copies to: c/o CFB Halifax, PO Box 99000 Stn Forces, Halifax, NS B3K5X5

1	ADAC(A) – Commanding Officer
1	CFMWC – Commanding Officer
1	CFNOS – Commanding Officer
1	MOG5 – Commanding Officer
1	PSU HALIFAX – Commanding Officer
1	TRINITY – Commanding Officer

International

1 Dr. James Pitton
ONR Global
Edison House
223 Old Marylebone Rd.
London NW1 5TH

1 Dr. Tor Knudsen
Director of Research, FFI
PO Box 115
N3191, Horten
Norway

LIST PART 2: External Distribution by DRDKIM (continued)
International (continued)

1 Dr. David Johnson
Office of Naval Research
875 North Randolph Street, Suite 1425
Arlington, VA 22203-1995
USA

1 Prof. Paul White
Institute of Sound and Vibration Research
University of Southampton,
Highfield, Southampton
S017 1BJ, UK

1 Prof. Victor Humphrey
Fluid Dynamics and Acoustics Group
Institute of Sound and Vibration Research
University of Southampton,
Highfield, Southampton
S017 1BJ, UK

1 Dr. Duncan Williams
Defence Science & Technology Laboratory, Physical Sciences
Winfrith Technology Centre
Dorchester, Dorset
DT2 8XJ, UK

1 Dr. David Liebing
National Leader, TTCP MAR TP-9
Defence Science & Technology Organisation, Maritime Operations Division
PO Box 1500
Edinburgh SA 5111
AUSTRALIA

1 Dr. Nathaniel de Lautour
National Leader, TTCP MAR TP-9
Defence Technology Agency
Naval Base, Private Bag 32901
Devonport, Auckland
NEW ZEALAND

1 Dr. Geoffrey A. Williams
National Leader, TTCP MAR TP-9
Defence Science & Technology Laboratory, Winfrith Technology Centre
Winfrith Newburgh
Dorchester, Dorset
DT2 8XJ, UK

1 Dr. Theo Kooij
National Leader, TTCP MAR TP-9
Program Officer, Littoral ASW
Office of Naval Research
875 North Randolph Street, Suite 1425
Arlington, VA 22203
USA

1 Dr. Keith L. Davidson
Undersea Signal Processing (321US)
Office of Naval Research
875 North Randolph Street, Suite 1425
Arlington, VA 22203-1995
USA

18 TOTAL LIST PART 2

39 TOTAL COPIES REQUIRED

This page intentionally left blank.

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. Centre sponsoring a contractor's report, or tasking agency, are entered in section 8.) DRDC Atlantic	2. SECURITY CLASSIFICATION (Overall security classification of the document including special warning terms if applicable.) UNCLASSIFIED

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

Aural Classification for Active Sonar: Final Report for TIF Project 11CQ11 "Aural Discrimination of True Targets from Geological Clutter"

4. AUTHORS (last name, followed by initials – ranks, titles, etc. not to be used)

Hines, Paul C.; Allen, Nancy; Young, Victor W.

5. DATE OF PUBLICATION (Month and year of publication of document.) May 2008	6a. NO. OF PAGES (Total containing information, including Annexes, Appendices, etc.) 34	6b. NO. OF REFS (Total cited in document.) 12

7. 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

8. SPONSORING ACTIVITY (The name of the department project office or laboratory sponsoring the research and development – include address.)

DRDC Atlantic

9a. 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.) TIF Project 11CQ11	9b. CONTRACT NO. (If appropriate, the applicable number under which the document was written.)
10a. ORIGINATOR'S DOCUMENT NUMBER (The official document number by which the document is identified by the originating activity. This number must be unique to this document.) DRDC Atlantic TR 2007-361	10b. OTHER DOCUMENT NO(s). (Any other numbers which may be assigned this document either by the originator or by the sponsor.)

11. DOCUMENT AVAILABILITY (Any limitations on further dissemination of the document, other than those imposed by security classification.)

Unlimited

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

Unlimited

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.)

Military sonars must detect, localize, and track submarine threats from distances safely outside their circle of attack. Active sonars operating at low frequencies are favoured for the long ranges they afford against quiet targets. However, in littoral environments, echoes from geological features (clutter) are frequently mistaken for targets of interest which results in degraded performance. Conventional signal processing techniques for dealing with this have met with only limited success *and* ignore a potentially valuable discrimination tool – the human auditory system. For it is generally accepted within the sonar community that operators can aurally discriminate target echoes from echoes from naturally occurring objects; however, the performance improvement has never been quantified. This report presents the results from a DRDC Technology Investment Fund (TIF) project whose aim was to measure the ability of the human operator to discriminate clutter from true echoes, to identify specific aural features that the operator employs, and to integrate these features into an automatic aural classifier. Using the area under the Receiver Operator Characteristic (ROC) curve, A_z, as a measure of performance, the results show that both the human operators and the automatic classifier ($A_z \approx 0.95$) score near to ideal performance ($A_z = 1$) for a broadband impulsive source. Performance degrades somewhat when acoustic bandwidth is reduced but is still very good ($A_z \approx 0.8$). Furthermore, results indicate that both the operators and the automatic classifier were performing the classification in essentially the same way. In addition to the target classification aspects of the project, DRDC collaborated with the Geological Survey of Canada (Atlantic) to identify high-clutter and low-clutter areas on the Scotian Shelf to support future trials that will examine active sonar performance in the littorals. As well as reviewing the project results, the future direction of the research is highlighted.

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 that 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.)

Audition, auralization, active sonar, ASW, DLC, classification, clutter discrimination

This page intentionally left blank.

Defence R&D Canada

Canada's leader in defence
and National Security
Science and Technology

R & D pour la défense Canada

Chef de file au Canada en matière
de science et de technologie pour
la défense et la sécurité nationale

DEFENCE DÉFENSE

www.drdc-rddc.gc.ca