AFRL-RH-WP-TR-2010-0007

Human Validation of the AUDIB Auditory Perception Model for Rotarywing Aircraft

Evelyn M. Hoglund
Nandini Iyer
Douglas S. Brungart
Frank S. Mobley
John A. Hall
Air Force Research Laboratory
Human Effectiveness Directorate
Wright-Patterson AFB OH 45433

Joseph Fernando
K Force
3085 Woodman Dr, #213
Dayton OH 45420

April 2008

Interim Report for August 2007 to March 2008

Air Force Research Laboratory
Human Effectiveness Directorate
711th Human Performance Wing
Warfighter Interface Division
Battlespace Acoustics Branch
Wright-Patterson AFB OH 45433

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report was cleared for public release by the 88th Air Base Wing Public Affairs Office and is available to the general public, including foreign nationals.

Qualified requestors may obtain copies of this report from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-RH-WP-TR-2010-0007 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

FOR THE DIRECTOR

//signed//
Robert McKinley
Program Manager
Battlespace Acoutstics Branch

//signed//
Michael A. Stropki
Chief, Warfighter Interfaces Division
Human Effectiveness Directorate
711th Human Performance Wing
Air Force Research Laboratory

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
2 Apr 2008	Interim	August 2007 – March 2008

4. TITLE AND SUBTITLE
Human Validation of the AUDIB Auditory Perception Model for Rotarywing Aircraft

5a. CONTRACT NUMBER
In-House

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
62202F

6. AUTHOR(S)
Evelyn M. Hoglund*
Nandini Iyer*
Douglas S. Brungart*
Frank S. Mobley*
John A. Hall*
Joseph Fernando**

5d. PROJECT NUMBER
7184

5e. TASK NUMBER
16

5f. WORK UNIT NUMBER
71841611

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
K Force**
3085 Woodman Dr, #213
Dayton OH 45420

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Materiel Command*
Air Force Research Laboratory
711th Human Performance Wing
Human Effectiveness Directorate
Warfighter Interface Division
Battlespace Acoustics Branch
Wright-Patterson AFB OH 45433-7901

10. SPONSOR/MONITOR'S ACRONYM(S)
711 HPW/RHCB

11. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFRL-RH-WP-TR-2010-0007

12. DISTRIBUTION AVAILABILITY STATEMENT
Approved for public release; Distribution is unlimited.

13. SUPPLEMENTARY NOTES
88ABW Cleared 06/26/2008; WPAFB-08-3864.

14. ABSTRACT

This study compares computational auditory detection model predictions against a corresponding large sample of human sound jury data points obtained in the laboratory. Helicopter and ambient soundscape signals were obtained from high sensitivity recordings in the field. Playback in the laboratory was achieved under high fidelity large volume headphones calibrated to accommodate helicopter primary rotor frequencies with minimal distortion above human sensation level. All sound jury members completed at least 12,000 trials detecting helicopters against wilderness, rural, suburban, and a variety of urban soundscapes, to represent the spectrum of potential environments involved in a real world scenario. Analysis compares the human sound jury performance against a contemporary computational auditory detection model, called "AUDIB", developed by the U.S. Army and NASA.

15. SUBJECT TERMS
auditory perception, auditory detection, auditory signal detection, auditory critical bands, helicopters

16. SECURITY CLASSIFICATION OF: Unclassified			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE		57	Robert McKinley
U	U	U	SAR		19b. TELEPONE NUMBER (*Include area code*)

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI-Std Z39-18

This page intentionally left blank.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

We would like to acknowledge the Tactical Technology Office (TTO) of the Defense Advanced Research Projects Agency (DARPA), in particular Mr. Dan Newman, for sponsoring this effort under Phase 1B of the Helicopter Quieting Program. We would also like to thank the Institute for Defense Analysis for assisting AFRL in the review of the literature pertaining to this effort. Collaboration with Dave Connor, Casey Burley and Kevin Shepard at the NASA Langley Research Center has been most valuable as well as the advice of Leonard Shaw, Dick Horonjeff and Dr. Fred Schmidt, each of whom has expertise in this scientific area spanning decades. Also, thanks to Jason Ross at HMMH for supplying the soundscapes of the Boston area, and Mark Ericson for his assistance on calibration. Special thanks to Kenneth Johnson, the lead audio/recording engineer for AFRL.

This page intentionally left blank.

EXECUTIVE SUMMARY

Abstract

Traditional auditory perceptual models for detection of complex signals against complex ambient soundscapes are based on the human audibility threshold imposed upon computed representations of auditory critical band filters. Such models attempt to locate a positive signal to noise ratio (SNR) in any singular band or group of bands and then apply classic signal detection theory to derive detectability measures (d prime, d′) and probability of detection (POD) values for the event. One limitation to these models is the low volume of experimental validation against real human sound jury performance, especially using very low frequency target signals such as helicopters. This study compares computational auditory detection model predictions against a corresponding large sample of human sound jury data points obtained in the laboratory. Helicopter and ambient soundscape signals were obtained from high sensitivity recordings in the field. Playback in the laboratory was achieved under high fidelity large volume headphones calibrated to accommodate helicopter primary rotor frequencies with minimal distortion above human sensation level. All sound jury members completed at least 12,000 trials detecting helicopters against wilderness, rural, suburban, and a variety of urban soundscapes, to represent the spectrum of potential environments involved in a real world scenario. Analysis compares the human sound jury performance against a contemporary computational auditory detection model, called "AUDIB", developed by the U.S. Army and NASA.

Introduction

Previous work related to auditory detection of U.S. military operations has resulted in computational models for predicting their audibility. However, these models have not been fully corroborated by studies in the laboratory using human listeners in time varying soundscapes. As such, the accuracy of the model involved has not been confirmed. The current study was conducted to validate one of the current auditory detection models (AUDIB), and to provide input regarding improvements for better prediction. The scope of this effort was limited to helicopters.

Background

Environmental Noise Research

Considerable research, with great success, has been conducted on annoyance, loudness scales, temporal summation and other perceptual metrics concerned with environmental consequences of helicopter and fixed wing aircraft noise on communities and on the wilderness. As a result, standardized metrics exist to describe and weight these effects (such as the “Noy” scale, the “Bark” scale, DNL, EPNL, SEL, etc.). However, most of these “environmental” noise metrics have little utility in predicting aural detection ranges for mission planning. These metrics are based on an A-weighted scale, which emphasizes sounds in higher frequencies, while many of the sounds related to aircraft detection are in lower frequencies. See Figure 1 for examples of signal-to-noise ratios (SNRs) for detection of aircraft in different backgrounds. The 117.4 mile camp background includes a greater level of high frequencies, clearly demonstrating that these metrics do not correlate with detection in all backgrounds. Nor can they adequately explain how humans use acoustics to classify and track aircraft in a dynamic context (Horonjeff, 2008). Additionally, these scales were developed to measure annoyance, rather than detection. While annoyance measures are based on the desire of the listener to NOT hear the signal, detection measures apply to listeners who DO want to hear the signal. As a result, the two kinds of scales measure different response biases for the same signal. An additional complication lies in the nature of the environment involved. The criterion for annoyance of the listener is adjusted according to the overall noise of the environment. Listeners in a city are likely to have a higher tolerance for detection of aircraft sounds than those in the national parks. Further, Fidell (1977) reported that below levels of about 65 dBA there is poor correlation between physical indices of exposure and annoyance judgements.

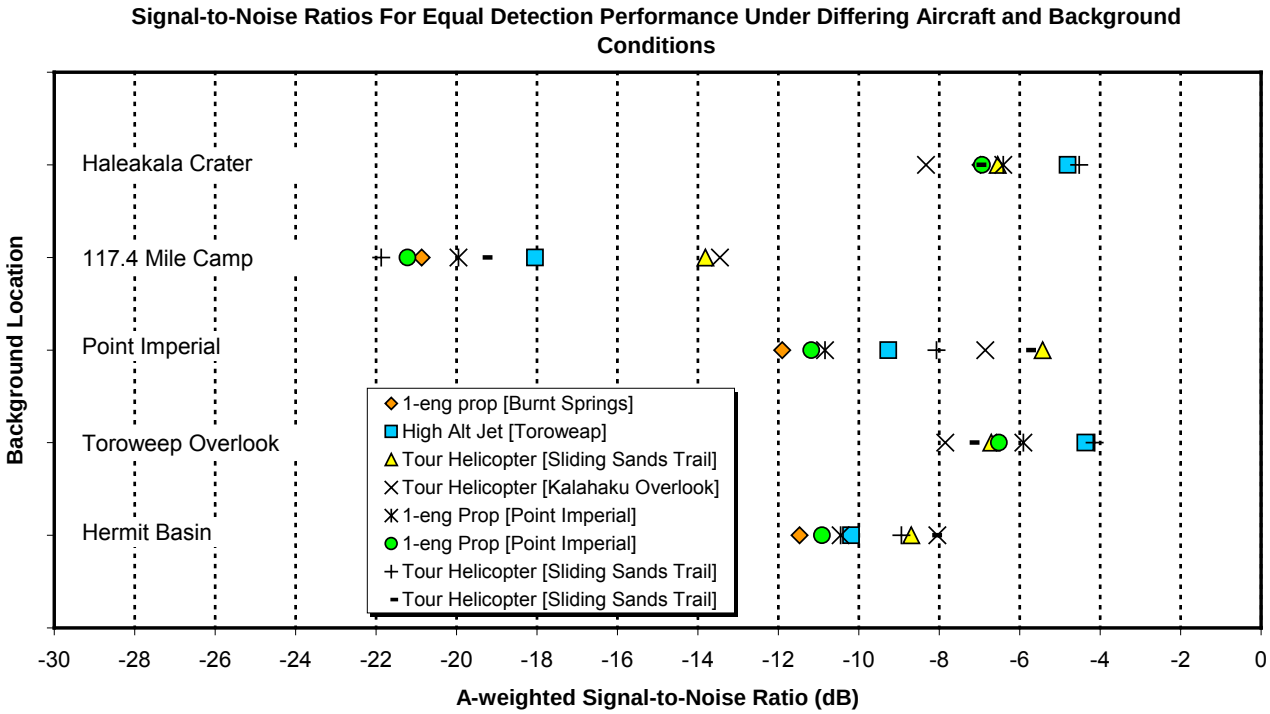


Figure 1: Detection SNRs for 8 aircraft in 5 background soundscapes (Horonjeff, 2008).

Helicopter noise generation

All vehicles have characteristic noise signatures, which allow them to be detected, identified, and classified by the human ear without additional technology. The primary sources of helicopter noise are the rotors and the engines, with three primary components. The first of these components is the rotational noise, which is caused by the differential air pressure from the blade passage and produces the helicopter's characteristic pulsatile sound. The second component is aerodynamic noise, produced by the disruption of the surrounding atmosphere caused by the helicopter, and is broadband in nature. The third component is the blade slap, which occurs only in some circumstances, such as during high speed flight or in maneuvers, and is caused by the blade passing through the vortex behind the previous blade of the main rotor (blade vortex interaction, BVI). Loewy (1973) specifically identified the primary noise sources as the engine on piston engine helicopters, and the rotors on turbine powered helicopters. Ungar (1972) provided a comprehensive summary of research defining the different aircraft components that contribute to the acoustic signatures of helicopters.

Human audibility/psychoacoustics

Critical Band Detection

Research into prior work and past experiments was conducted independently by AFRL and the Institute for Defense Analysis. Both teams concluded the bulk of meaningful attempts to build predictive algorithms for aural detection are based on some implementation of auditory critical band filters (Ollerhead, 1971). These critical band functions were first described by Fletcher (1940), and later by Zwicker, Flottorp, and Stevens in 1957. Through a series of psychophysical experiments, they developed a set of frequency-based filters that correspond to the frequency resolution of the human auditory system. Later work by other researchers has further described the width of these critical bands (Greenwood, 1961, Moore and Glasberg, 1983, among others). The Moore and Glasberg (1983) calculation for the critical band is referred to as the equivalent rectangular band (ERB), as it is determined to be the width of the width of a bandpass filter with infinitely steep slopes, thus forming a theoretical rectangular filter. Essentially the models attempt to determine if, relative to the sensitivity (threshold) of human hearing in each critical band, there is sufficient target signal relative to the background ambient noise, to trigger detection. This construct is then coupled to classic signal detection theory (Green, 1959) to produce a Probability of Detection (POD), and/or d', for each time step in the model. Favorable PODs are looped back through sound propagation calculations to predict the far field range at which the aircraft would be detected. Further evidence for the applicability of signal detection was reported by Fidell, Pearson, and Bennett (1974), when they compared a statistical prediction model with a d'_{max}. Their results indicate that the signal detection predictions were a closer match to the empirical results than the statistical predictions.

Theory of Signal Detection (TSD)

The theory of signal detection (TSD) describes the performance of an ideal observer in the detection of signals in noise. This allows for the separation of the sensitivity of the observer from other components of the decision process, e.g. response bias or internal "noise" such as memory or attention. TSD uses statistical methods to calculate the performance of the ideal observer on the basis of comparison between a distribution of noise alone, and a distribution of noise with a signal. Each distribution includes the range of possible variations in the waveform to be detected. Thus, a decision regarding detection of a signal may be classified in one of four ways: positive responses may be correct (if from the signal + noise distribution, a hit) or incorrect (if from the noise only

distribution, a false alarm), while negative responses may be correct (if the sample is from the noise only distribution, a correct rejection) or incorrect (from the signal + noise distribution, a miss). The detection measure d' (d prime) is based on a normalized distance between the means of the two distributions, with d' = 1 being equivalent to one standard deviation. [Thus, for a d' = 1, the means are separated by one standard deviation, for a d' = 2, the means are separated by two standard deviations, etc.] One value of the d' measure is to account for differences in response bias of the observer. The response bias exhibited by a human observer will affect the proportions of 'yes' and 'no' responses to the experimental signals, but the d' measurement is independent of the bias. The response bias depends on the probability that a signal will occur, and on the relative rewards for correct responses, versus the cost of incorrect responses. In a tactical situation, the cost of missing a signal could be loss of life, whereas identifying a signal that is not actually there may simply be excess use of ammunition. In this scenario, the response bias would be in favor of 'yes', but the d' may remain unchanged relative to a different cost/benefit ratio.

Aircraft aural detection/classification

A number of studies have been conducted under the sponsorship of the U.S. military to quantify the aural detection of aircraft. Among these are studies are projects measuring detection in field conditions. A study by Hartman and Sternfeld (1973) tested the model presented by Ollerhead (1971), which was developed in the laboratory, in a field study. They found the model's detection prediction to be extremely conservative, both when analyzed by sound pressure level (SPL) of the acoustic signal and by distance of the helicopter from the subjects. That is, the subjects did not detect the helicopter until a much higher level relative to the ambient, and at twice the distance the model predicted. They offer a possible explanation of the difference as their study being the more representative, but less critical, model for aural detection.

A study reported by Abrahamson (1975), also analyzed helicopter sound propagation and human aural detection in a field environment. His subjects were to indicate both when they thought they heard a helicopter, and then again when they could confirm that they heard it. One group was to focus on the listening task, while another group was given other tasks as diversions. The results indicate that the first responses (uncertain detection) appear to be based on low frequency components, while the late responses (certain detection) are based on higher frequency components. This study confirmed the results reported by Ollerhead that showed that helicopter signals could be masked at 5 dB below the ambient critical band spectrum level.

Similarly, in a review of available data, Loewy (1973) concluded that based on factors related to ambient sound conditions and terrain, auditory

detection is due primarily to components in the first three octave bands of the sound. He also concluded that components above 300 Hz could be detected as low as 9 dB below the ambient, with components below that dependent primarily on limits in the human auditory response. His conclusions were not based on human detection results, however, as his analysis was focused on the goal of helicopter noise reduction.

Ungar et al. (1972) reviewed a wide range of studies related to helicopter noise generation and the effects of different aircraft components on the noise signatures. In his report, appendix N addresses briefly the issue of auditory detection, with an overview of the reported masking effects of different environments, particularly jungles and forested areas. His summary indicated that detection levels increased with increasing density of vegetation. He further reported that detection was better at night in the low frequencies, but better for the high frequencies in the daytime. His overall conclusion from review of existing data was that the lowest levels for detection were at midday (easiest detection of the helicopters) and the highest levels were in the early evening hours (poorest detection).

A number of researchers have also developed models for the auditory detection of aircraft by human listeners. Some of these include Taylor and Poe (1973) and Elshafei, Akhtar, and Ahmed (2000), and the AUDIB model produced by Wyle Labs, beginning in 1975 as the I Can Hear It Now (ICHIN) model, developed for the U.S. Army. One of the difficulties presented by all of the models reported is a lack of corroboration by empirical data from human listeners.

The field studies described above rely on real world conditions. This is both a conceptual strength and an experimental design weakness. The strength assumes no doubt about the realism of the target signal, because the signal is live. However, the weakness of the approach is found in atmospheric and aircraft states which can vary across trials, thus confounding the reliable duplication of the signal at the listener across multiple trials. The signals, by being live, all include both the target aircraft and the background environment, making it impossible to separate the two components and analyze the SNR. These variables also make it impossible to determine what factor is most responsible for detection, whether a part of the signal or a variation in the background.

Horonjeff, Fidell, and Green (1983) reported a series of experiments using laboratory created signals to measure specific factors in detection of periodic impulse sounds, as a more critical measure for auditory detection thresholds that would relate to aircraft such as helicopters. In this study, they measured detection thresholds for impulses at repetition rates in the range of helicopter rotor frequencies. Their overall conclusions for their signals were that the individual pulses summed in a predictable manner for detection, and that this

summation is “leaky”, that is, greater signal energy is required for detection with longer observations.

The strength of the AFRL conducted experiment described below is in the ability to exactly duplicate the target versus ambient noise in the trials presented to the sound jury subjects, and to randomize the presentation intervals in order to maximize the statistical power of the experiment. At the same time, real helicopter signals are modified and used, rather than simplified laboratory generated signals. Furthermore, this experiment isolated the acoustic characteristics of the signal and the noise at the listener from acoustic variance in the source or the propagation of the signals. The listener judgments were purposefully decoupled from non-psychoacoustic factors, such as attention. Each sound jury subject was presented with at least 12,000 intervals of target versus ambient signals. By controlling for confounding variables that can be introduced by issues with calibration, recording quality, playback quality, headphone response, and listener state, this study could ensure high confidence in the experimental results.

Methods

Experiment Description

Subjects

Fourteen members of a panel of paid subjects, ranging in age from 19 to 57 years with normal hearing acuity, participated in the psychoacoustic (human detection) portion of the study. Normal hearing was defined as air conduction thresholds at 20 dB HL or better for octave frequencies between 250-8000 Hz. Each subject’s hearing was retested on a regular basis to ensure continued qualification for studies with the requirement for normal hearing. All subjects were well trained for psychoacoustic experiments, with prior experience in other auditory studies in this laboratory.

Hardware and Software

Sound Recordings

The acquisition of the acoustic information for this study was accomplished by Harris, Miller, Miller, and Hanson (HMMH) of Burlington, MA and AFRL. The efforts accomplished by AFRL will be described here. This data represents the stimuli and six of the nine ambient waveforms. The data acquisition for the helicopter stimuli was accomplished at Eglin Air Force Base. The recording equipment was comprised of three G.R.A.S. low noise microphone systems. Each system has a microphone power supply (G.R.A.S. Type 12HF), preamp and microphone that are matched by the manufacturer. Three microphones were used in the measurement system. One was in the free-field at four feet above ground level to obtain the monaural recordings, and for use as a reference microphone. The reference system was the Type 40HH, Figure 2. Two other microphone systems of Type 40HT, Figure 3, were also used to capture binaural recordings within a Knowles Electronics Mannequin for Acoustic Research (KEMAR®). The KEMAR®, Figure 4, is a head and torso simulator (HATS) which meets the requirements of ANSI S3.36/ASA58-1985. Both microphone systems were arranged in close proximity to each other with a burlap wind screen as shown in Figure 7.

Figure 2: G.R.A.S. 40HH low noise system.

Figure 3: G.R.A.S. 40HT low noise system.

Figure 4: KEMAR mannequin.

The signal acquisition for the low noise microphones was accomplished through the use of a CF-18 Panasonic Toughbook® and a National Instruments cDAQ-9172 CompaqDAQ chassis. The chassis was loaded with the NI-9211 24-bit 50,000 samples/sec sample rate DAQ boards. The data was collected through the use of a customized interface built on top of the National Instruments DAQmx technology. Each of the microphone outputs was stored in a 32-bit floating point mono canonical wave file.

Analysis for the wave files that were presented to the subject was accomplished through use of the National Instruments LabVIEW® Sound and Vibration toolkit. The toolkit implements one-third octave band filters that are compliant to the ANSI standard. The desire was to have fractional octave outputs from 10 Hz to 16,000 Hz for each of the ambient and stimulus files. The stimulus files were 1 second in duration. The settle time for the filterbank in the analysis due to the 10 Hz low frequency is 2.5 seconds. To compensate for this, the one second waveform was concatenated with itself three times to create a 4 second long waveform. The ambient files were analyzed with a 1 and 0.5 second integration time to achieve the levels at the 0.5→1.5 and 2→3 second intervals. These time samples correspond to the stimulus intervals in the experimental procedure used with the human listener data collection.

Headphones

Headphones for presentation of the auditory signals were selected on the basis of the response in the low frequencies. Headphone response curves can be found in Appendix 2, for presented frequencies at 10, 20, 30, and 63 Hz. The BeyerDynamic DT-990 headphones were chosen because, of the available headphones, they demonstrated the least harmonic distortion in the low frequencies. The greatest amount of distortion was found with the 10 Hz and 20 Hz tones, with harmonics between 500 and 1000 Hz at up to 20 dB SPL above the human audibility curve. This is a low level of distortion, and all of the ambient soundscape levels were above this, thus this distortion was a minimal concern for this study.

Stimuli

All sound stimuli were digitally manipulated using Adobe Audition® and MATLAB® for presentation to subjects. The recordings used a 48000 samples/sec sampling rate, and 24-bit digitization stored in 32-bit form for the amplitude. The stimuli were then presented with 16-bit digital to analog conversion through MATLAB®. Stimuli are divided into target and ambient categories, with targets defined as the auditory signals to be detected, and ambients defined as the noise backgrounds in which the targets are presented.

Targets:

Target stimuli consisted of 1-second samples of helicopter signals taken from recordings made at the Eglin Air Force Base in August and September of 2007 during a military program known as “Chicken Little”. These target signals were obtained by selecting portions of recordings that included the approach and near flight of two different helicopters (MD-902 and MI-8). Portions of the recordings that included departure were excluded from the study, as the purpose for the study was for detection of approaching aircraft.

Recordings for the MD-902, Figure 5, aircraft were made on the mornings of 23 and 24 August 2007. Recordings for the MI-8 aircraft were made on the mornings of 8 and 9 September 2007. Monaural and binaural detection results were obtained using 175 exemplars for the MD-902 helicopter, and 236 exemplars for the MI-8 helicopter (Figure 6), for a total of 411 targets. Exemplars were defined as discrete 1 second samples of the recordings, from which the target signals were selected.

Figure 5: MD-902 helicopter

Figure 6: MI-8 helicopter

Foils:

Within the experiments, foils were used to provide a signal without a helicopter in the reference interval. By introducing sounds taken from a recording from the same environment in which helicopter signals were taken, increased confidence can be achieved that the target signal is being detected on the basis of the helicopter present in the target recording, rather than other spectral components related to the ambient soundscape during the helicopter recordings, since those components are presented in both intervals. These signals were taken from one of the Eglin recordings that included primarily insect sounds.

Ambient soundscapes

Eglin

Ambient noise was obtained from the recordings made on 8 and 9 September in an open field on Eglin AFB, Florida (Figure 7). The recordings were made in the early morning between "Chicken Little" flight tests. Samples were 5 minutes in length and selected from portions of the recordings in which no helicopters could be detected. Out of a total of 28 such samples three were used in the study as ambient sounds. The three signals used as the ambients were selected to represent the quietest of the recordings, the loudest, and a midpoint level. The quietest sample (Ambient 19) included no discernable environmental noises, the midpoint sample (Ambient 5) included primarily insect noises and occasional birds, and the loudest sample (Ambient 28) included sounds of clothing rustling and some speech and other human generated sounds. Due to technical difficulty with the recordings with the KEMAR®, these were not included in the binaural portion of the study.

Figure 7: Eglin AFB recording setting

Boston

Three recordings were obtained from Harris, Miller, Miller, and Hanson (HMMH), an acoustics consulting firm located in Burlington, MA. These recordings were made in an urban park (Boston Common), a suburban street (Newton, MA), and a rural road (Boxford, MA), Figure 8-10. The ambient signals were 5-minute selections extracted from these recordings that were consistent for content and representative of the overall environment. The urban ambient soundscape included a variety of traffic noises including trucks, back up signals, and sirens recorded from in the park. The suburban soundscape included automobile traffic, birds, and pedestrians. The rural soundscape included birds and insects, as well as occasional distant ground and air vehicles. These recordings were only available for monaural signals, and, as such, were excluded in the binaural portion of the study.

Figure 8: Boston urban recording setting

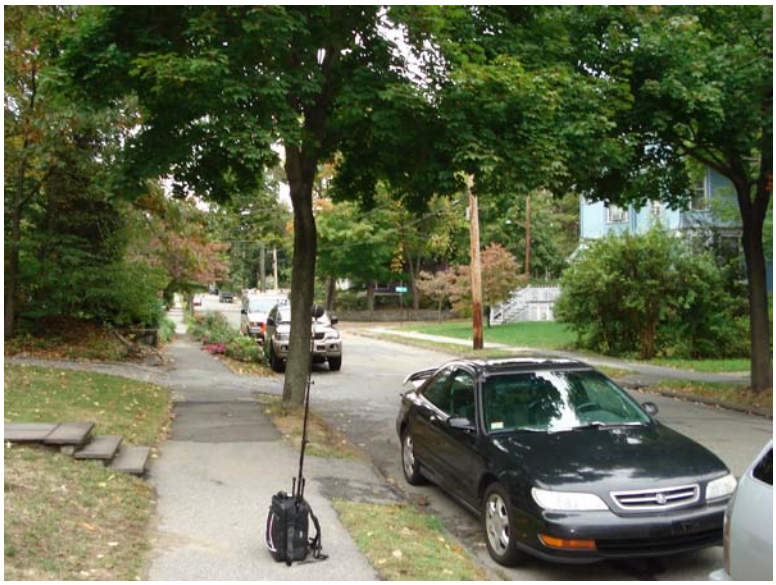


Figure 9: Boston suburban recording setting

Figure 10: Boston rural recording setting

Downtown Dayton

Three recordings were also obtained in Dayton, Ohio, to provide additional soundscapes. These all included urban settings, but with different environmental characteristics. Recordings were made at midafternoon at a downtown intersection surrounded by tall urban buildings, in front of the city courthouse (Figure 11), which was elevated from street level and across from tall urban buildings (Figure 12), and near the ATM at the entrance to a bank (Figure 13), with acoustic characteristics representative of an urban canyon (multiple reflective surfaces). These samples consisted of various traffic noises and

speech, with differences in the environments consisting of the number of reverberant surfaces, distance from traffic, and elevation. All recordings were made in close proximity to the noise sources. Five-minute selections from these samples were taken, based on overall consistency in the components of the soundscape, as well as the overall level.

Figure 11: Dayton recording setting – courthouse

Figure 12: Dayton recording setting – 3rd Street and Patterson

Figure 13: Dayton recording setting – National City Bank

Descriptions and spectrograms of the specific soundscapes used can be found in Appendix 1, Table 1-1 and Figure 1-1.

Method

Human Detection

Training

Prior to beginning the experiments, all subjects were provided several days of training on the task to assure that they were familiar with the target and ambient signals, as well as the overall task. Conditions during the training period were identical to those used during the monaural experiment, described below.

Monaural

A two alternative forced choice (2AFC) procedure was used, with 50 trials in each run. Signals were presented diotically, that is, the same signal was presented to both ears, so that the perception was centered between both ears.

Experimental runs were produced by randomly selecting and playing a 5-minute ambient sound from among the nine possible alternatives. Within an experimental run, a trial consisted of a 500 msec preparation interval, followed by a 1 second stimulus interval, a 500 msec interstimulus interval, another 1 second stimulus interval, and a 3 second response interval. Thus, each trial was 6 seconds long, illustrated in Figure 14. Within each trial, the target signal was randomly presented in either the first or the second stimulus interval. Subjects were asked to indicate which stimulus interval contained the target signal within the ongoing ambient soundscape, and were instructed that the target signal was one of the helicopters indicated on the response screen (Figure 15). The alternate interval contained a foil, consisting of a one second sample taken from an ambient soundscape recorded on the Eglin range. Target signals and ambient soundscapes were scaled to represent the actual relative intensities in the field, to compensate for level differences introduced by the equipment used for presentation. The scaling factor was established by comparing the output from the laboratory equipment to a 94 dB calibration tone recorded at the same session as the signals. For each trial, the target was played in one interval, scaled to 0, -10, -20, or -30 dB relative to the ambient. The RMS amplitude for each 1-second target ranged from -53 dBv to -18 dBv, and the RMS for the ambient soundscapes ranged from -54 dBv to -32 dBv. Resulting actual signal-to-noise ratios (SNRs) ranged from -60 to 30 dB for specific trials. The foil was adjusted with the target signals to equalize for the ambient components included in the target recordings. Subjects used a computer mouse to select the interval in which they heard the helicopter, and which aircraft it was. In this way, data related to detection and classification could be obtained simultaneously. Feedback was provided following every trial. Subjects continued with the data collection until a minimum of 12,000 trials were completed.

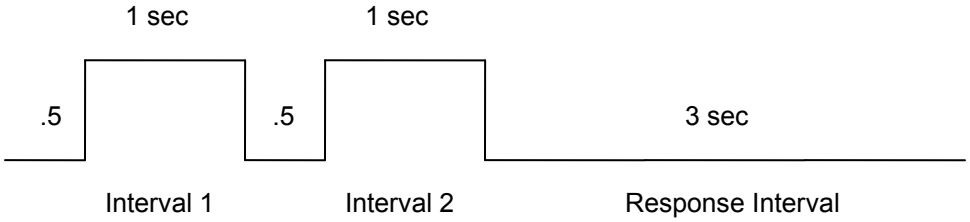


Figure 14: Diagram of an experimental trial.

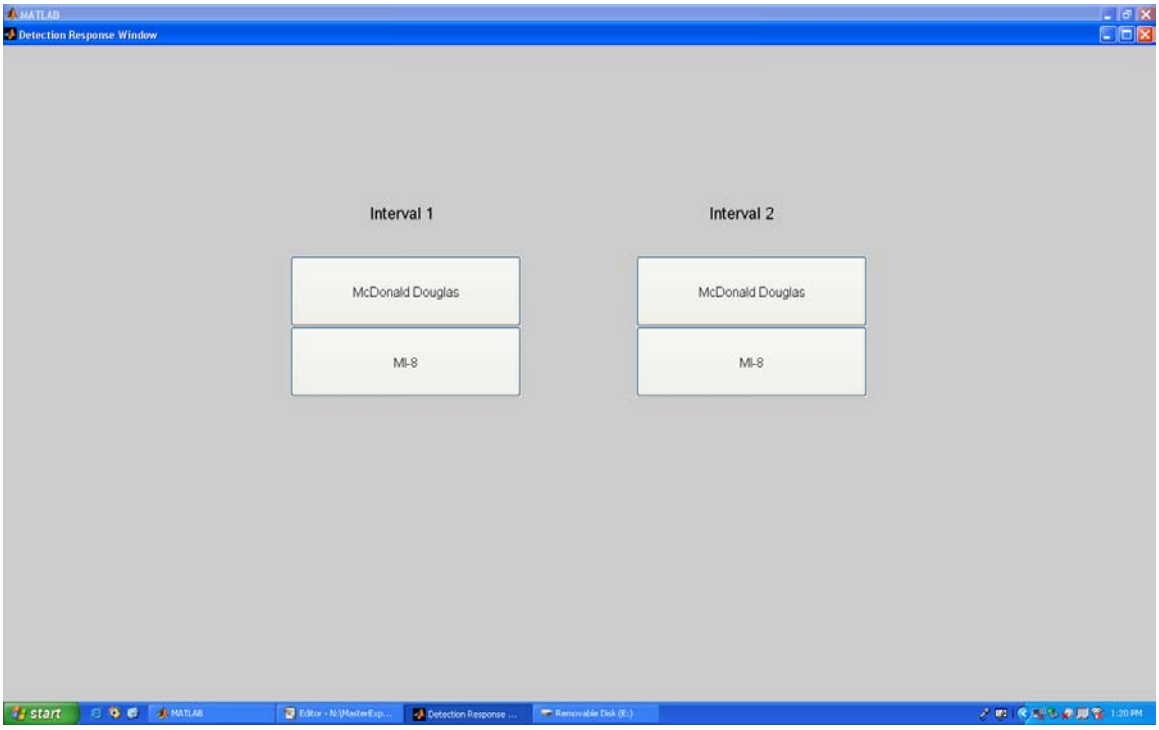


Figure 15: Screenshot of computer response screen.

Binaural

The same 2AFC procedure was used for the binaural detection study. The signals and ambient soundscapes were matched to those used in the monaural study, using recordings from the microphones installed in KEMAR®. This provides a representation of the effect of an average human head on a signal. Subjects again were asked to select the interval containing the target signal, with actual SNRs ranging from -60 to 30 dB relative to the specific ambient soundscape. Subjects collected a minimum of 12,000 trials for this study, as well.

AUDIB model

To test AUDIB against the human subject data the FORTRAN source code was compiled and used through the MATLAB® interface to run the application. Routines to write the case file and the associated data files were written in MATLAB®. A basic description of the AUDIB functions can be found in Appendix 3. For this first execution of the model each of the stimuli were compared to a 'long-term' ambient spectrum. The five minute ambient files were run through MATLAB®'s built-in FFT function. The resolution of this FFT was 48 Hz. The one second target signal that was input to AUDIB was a FFT with the

same frequency resolution. A modification was then made to the AUDIB model in which the ambient levels were presented to AUDIB in time samples corresponding to the experimental intervals. This was called the 'short-term' data. A further adjustment was made by converting the FORTRAN code to MATLAB®. Analysis was limited to the lower frequency bands.

Data Analysis

Human detection

In all figures, data for -60 dB SPL SNR and for probability of <0.4 have been edited due to limited exemplars in these data. As a result, the reliability and validity of these data points is limited, and were excluded from further analysis.

Overall Results

Overall results from the human listener panel for the monaural detection study are shown in Figure 16. In this figure, the results are plotted only for the target amplitudes, using the RMS power (dB SPL) of the one second target signal for the measure. The data have been binned together into 6 dB wide bins and averaged together to obtain the individual data points shown in the figure. Two features can be seen from this figure. First, it is apparent that the probability of target detection increases with the overall level of the target signal (as indicated by the general increase in the curves from left to right on the x-axis). Second, it is clear that, for any given target level, the probability of detection systematically decreased as the ambient soundscape level increased from a relatively quiet environment (those collected at Eglin AFB and the Boston Rural and Suburban soundscapes) to a relatively loud environment (the Dayton and Boston Urban soundscapes). Notably, these appear to show a clear distinction between rural/suburban settings and urban settings. For example, the data appear to show a higher average level of detection performance in the Boston Suburban environment with a mean level of 43 dB SPL than they do for the Boston Urban environment with a mean target level of 73 dB. This suggests that the effective masking level of the urban environment was more than 30 dB higher than that of the suburban environment, compared with a difference of only 14 dB in the A-weighted dB SPL (L90) of the urban and suburban environments (56 dB versus 39.8 dB SPL). As discussed later, this may suggest that the kinds of sounds present in the urban environments (engine sounds, etc.) were more similar to the target helicopter sounds than those present in the more rural environments, and thus listeners had a much harder time identifying the helicopter sounds in the urban environments.

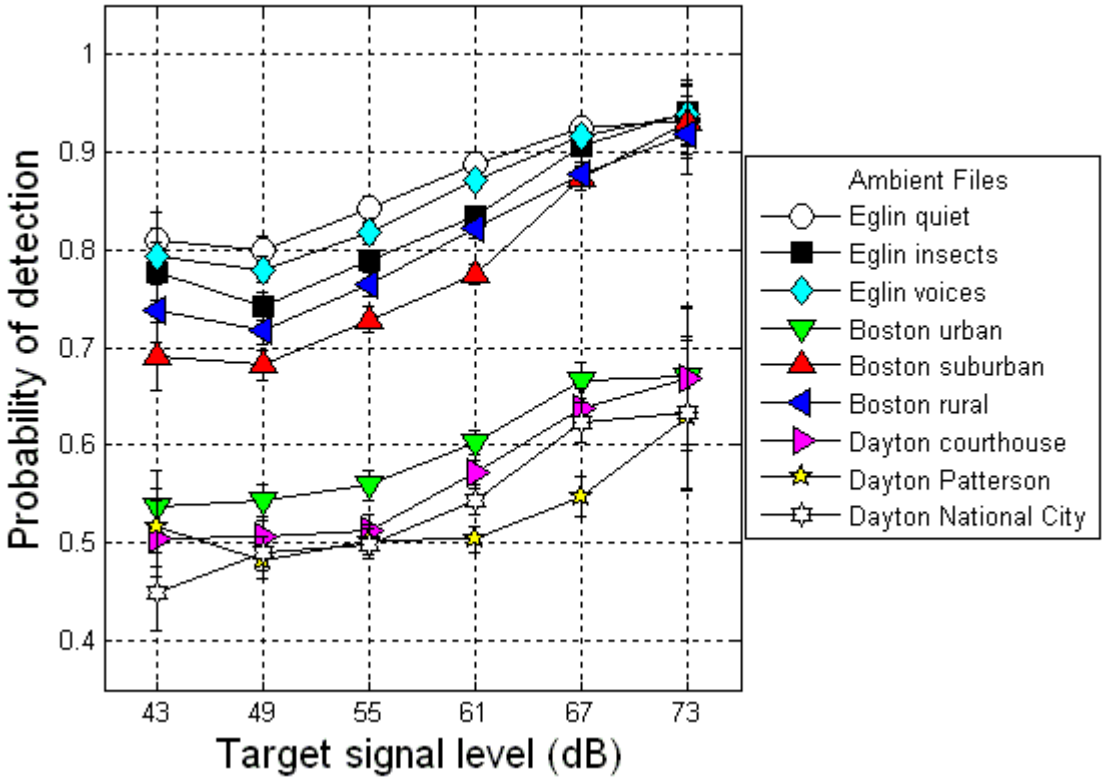


Figure 16: Human detection results plotted by target amplitude.

Analysis by SNR

The data plotted in Figure 16 do not account for the instantaneous variations in the level of each ambient soundscape across the different 1 sec target intervals for that ambient condition. In order to collapse across different soundscapes in a meaningful way, a better strategy is to calculate the total target energy and total masker energy in each stimulus interval and determine the instantaneous SNR for each individual trial in the experiment. This instantaneous SNR value was calculated by comparing the RMS amplitude of the signal with the RMS of the ambient. For example, the data point for -20 dB SPL SNR includes all responses to targets with a SNR between -15 and -25 dB SPL. The number of trials represented in each bin is dependent on the level of the target signal as well as random variations in the ambient levels, which were not controlled, thus some of the bins have a limited number of trials. Figure 17 shows the average probability of detection for human listeners plotted as a function of SNR for the monaural signals. These plots exhibit a very different profile of detection for the very quiet ambient soundscapes of the Eglin recordings than for the other ambient soundscapes used. Probability of detection increases rapidly between -50 and -10 dB SPL SNR, where it reaches ceiling, and all target sounds are detected with occasional errors incidental to the

procedure. In part, at least, the very high performance levels obtained with the Eglin recordings may reflect the inclusion of low frequency wind noise in the RMS estimates of total masker power, which may have inflated the apparent overall level of performance in these conditions.

The Boston and Dayton ambient soundscapes include sounds common to more populated areas, and result in probability of detection that does not exhibit improvement until the SNR reaches -30 to 10 dB SPL. In the ranges from -10 to 0 dB SPL SNR, the probability of detection for the signals decreases across ambients that increase in human generated sounds, such as traffic sounds. Specifically, in the rural soundscape, probability of detection is better than suburban, which, in turn, is better than the urban soundscapes (with detection the poorest in the Dayton ambients). These results are consistent with the current understanding of human detection thresholds for target sounds in noise.

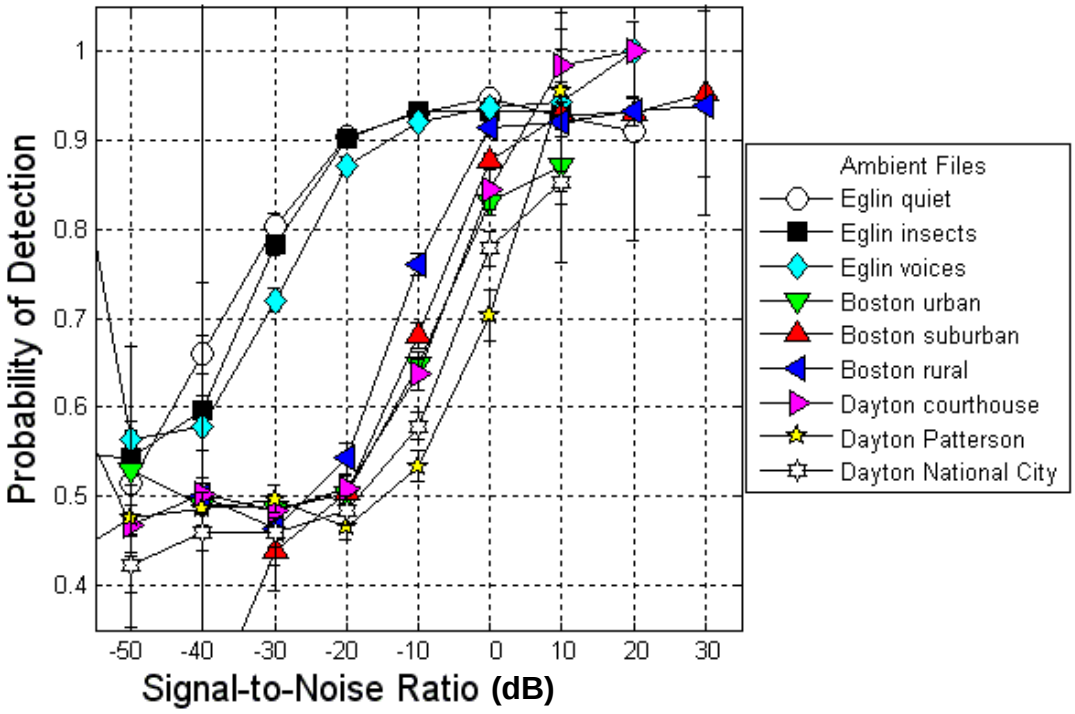


Figure 17: Probability of detection for human listeners with monaural signals. Data points for -60 dB SPL SNR and for probability of <.4 edited from figure due to limited exemplars in these data.

Binaural

Detection for the binaural targets included only the Dayton urban soundscapes due to availability of binaural recordings for only these soundscapes. The results of this study are shown in Figure 18. The results are consistent with the monaural results, with an improvement in detection between SNRs of -30 and 10 dB SPL. An improvement in the detection performance can be seen in these plots relative to the monaural detection.

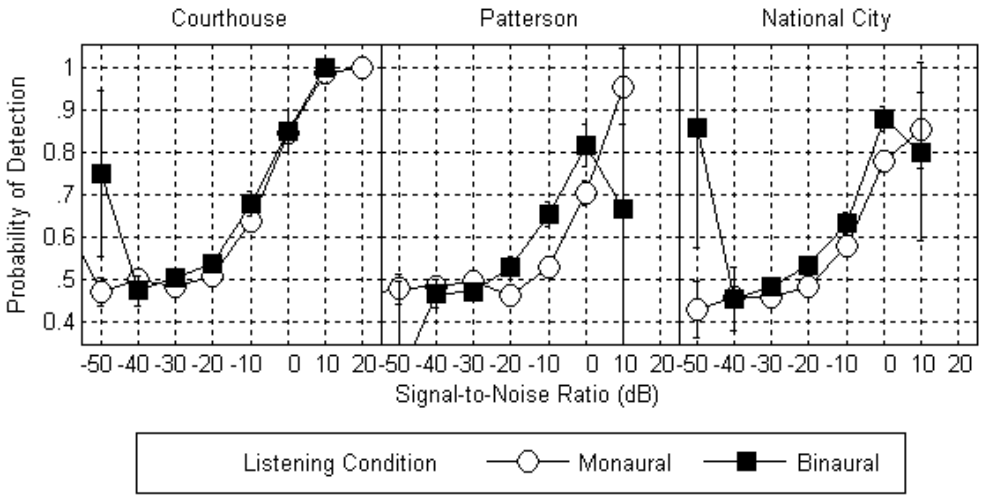


Figure 18: Probability of detection for human listeners for monaural vs. binaural signals. Data points for -60 dB SNR and for probability of <.4 edited from figure due to limited exemplars in these data.

AUDIB model prediction

Even if the results from the Eglin ambient soundfields are eliminated, the results in Figure 18 show that overall, flat-weighted SNR is not a great predictor of human detection performance. In fact, the SNR value required for 70% correct detection varied about 15 dB across the six non-Eglin soundscapes. In order to obtain a better estimate of human detection performance, a more sophisticated model that accounts for the detection of the stimulus in different frequency bands is necessary. Thus, the data were also processed with the AUDIB model.

Processing of the same target and ambient sounds through the AUDIB model yielded the probabilities of detection displayed in Figure 19. The model predicts an increase in detection at lower SNRs in quieter ambient soundscapes, with higher SNRs required for detection of the target signal as the ambients increase in overall content. Thus, more signal energy is required for the target to be detected when the ambient noise is denser. This pattern does not appear to be consistent, however, as the model predicts greater detectability in the National City Bank ambient (with the greatest amount of reverberation) than in the other

soundscapes from populated areas with less spectrally dense envelopes. The detectability predicted in this environment is equivalent with that in the loudest Eglin environment, which consisted primarily of voices and rustling sounds. The acoustic content of these two ambients is very different, yet the model predicts very similar results. Additionally, the probability of detection for the signal in the Dayton courthouse ambient is much lower, although the acoustic content of this environment is very similar to the Dayton Patterson environment. In general, the model predicts comparable detection for the quietest environments, and comparable detection for most of the moderately dense environments. Particularly for the more populated environments, this does not account for variation in the actual acoustic environments. The grouping of the probability curves cannot be easily explained on the basis of acoustic information in the ambients, other than the overall trend is for detection to require a greater SNR with increased noise density.

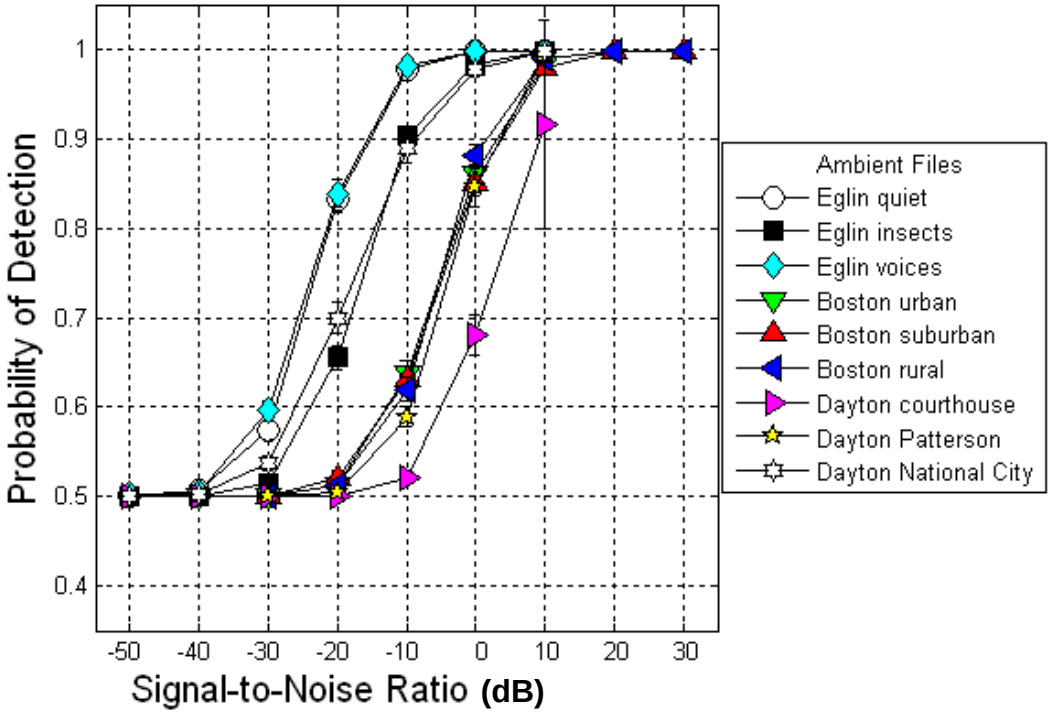


Figure 19: Model data from AUDIB (with long term integration). Curves reflect detection of 1 second targets predicted in the context of the overall average level of the ambient over a 5 minute sample.

Comparison of human and AUDIB results as a function of SNR

The human panel data and AUDIB predictions are represented together by ambient soundscapes in Figure 20. When plotted by SNR, the AUDIB model provides a good prediction of target detectability for the ambient soundscapes with a moderate noise level. These include the Boston recordings, the loudest of the Eglin recordings, and the Dayton courthouse and Patterson recordings. For the very quiet ambient soundscapes the model predicts poorer detection, while in the loudest soundscape, the model predicts better detection than the human results.

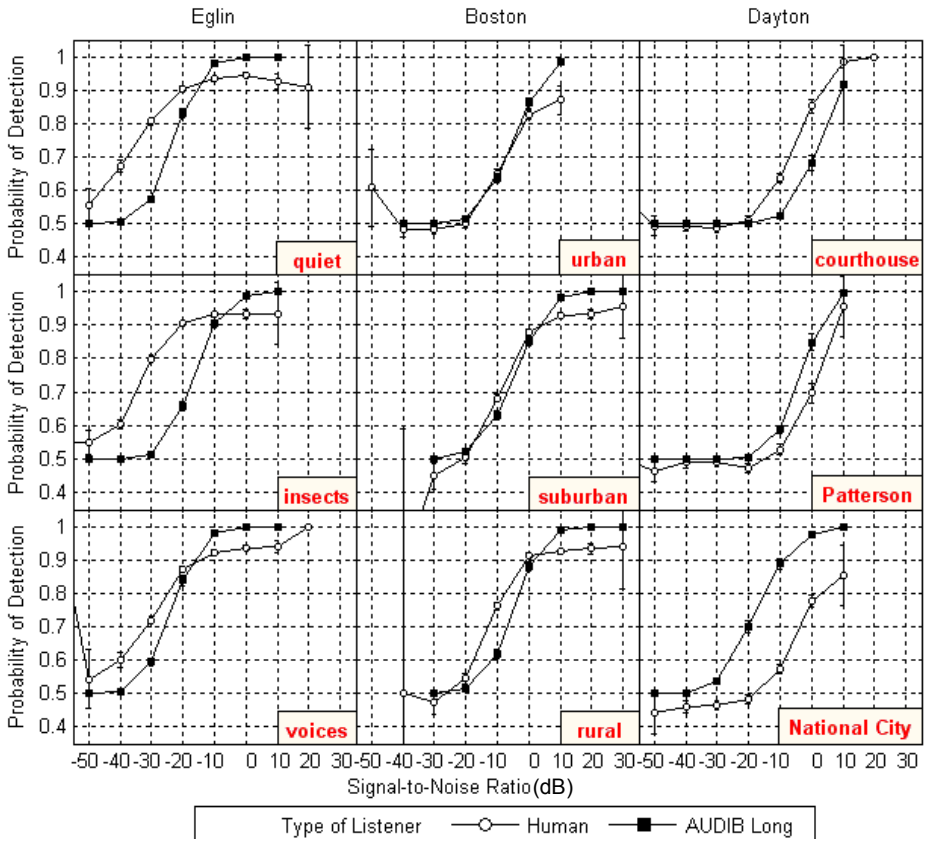
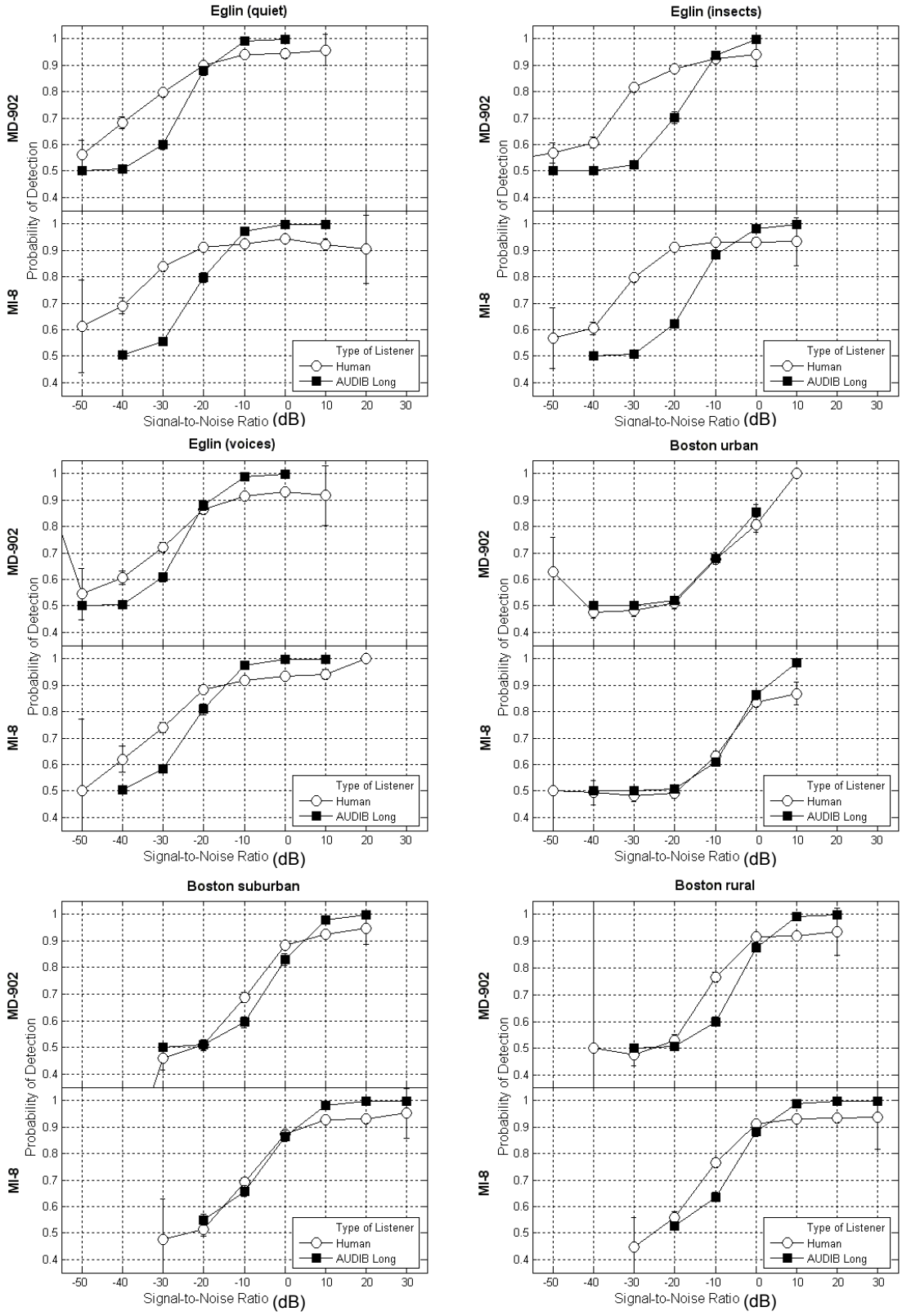


Figure 20: Comparison of current AUDIB model results with human data based on SNR.

Although some differences can be found on the basis of the type of aircraft to be detected, the differences in probability of detection between the model prediction and the human results are maintained when the data are analyzed for each type of target individually, as can be seen in Figure 21.



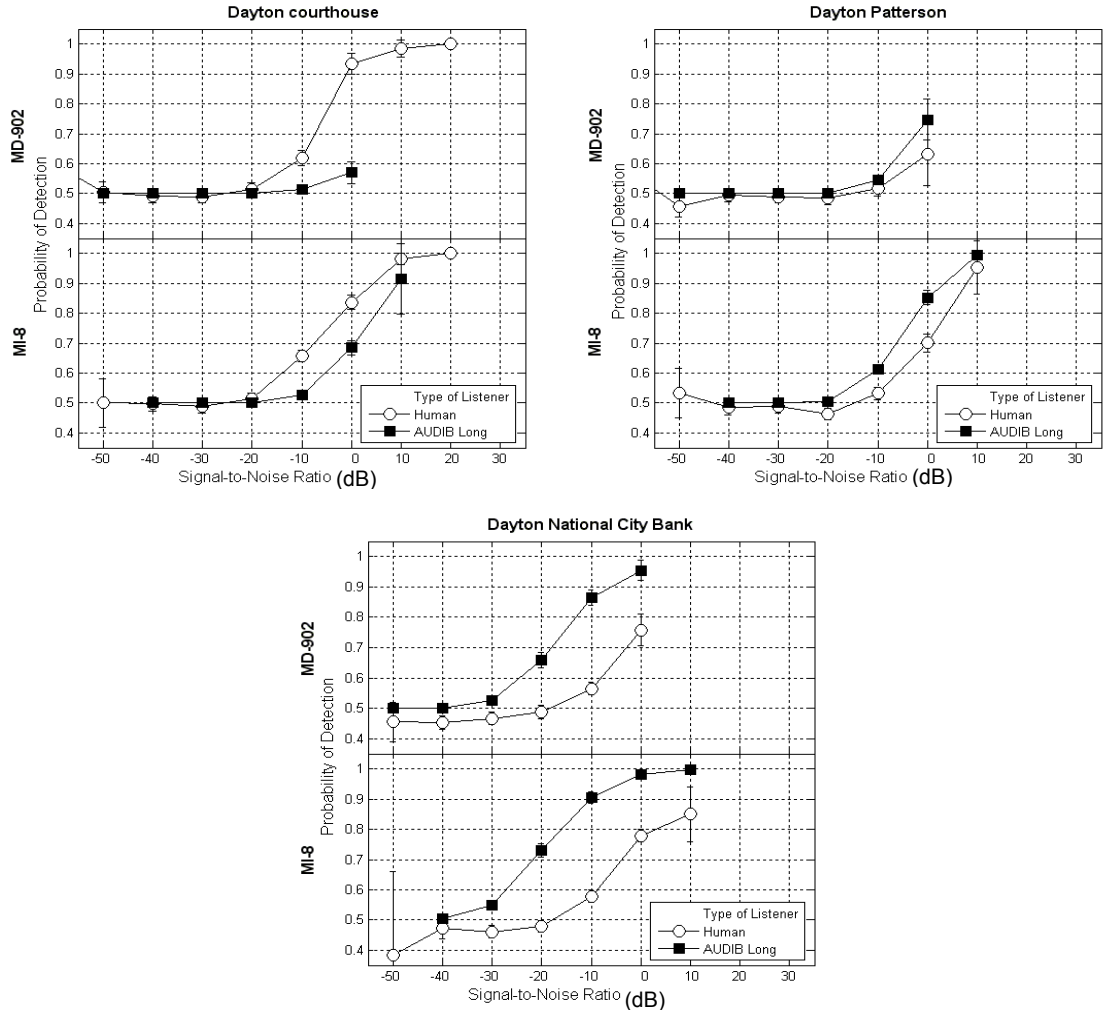


Figure 21: Comparison of human and AUDIB results for each type of helicopter in each soundscape, based on SNR.

While the AUDIB model in its current form provides good prediction for environments with some ambient noise, it does not hold up well for environments at the extremes, either quiet or loud. In these cases, it underpredicts or overpredicts auditory detection, respectively. As a result of the model's limitations in ability to match the human results, a modification of the model was implemented, in which the ambient noise was averaged over the same 1 second time sample as the target.

Classification

The data collected from the human subjects included classification of the two helicopters. These results can be seen in Figure 22. As with detection, the listeners are able to classify the helicopters with approximately equal accuracy, increasing with SNR in all ambients.

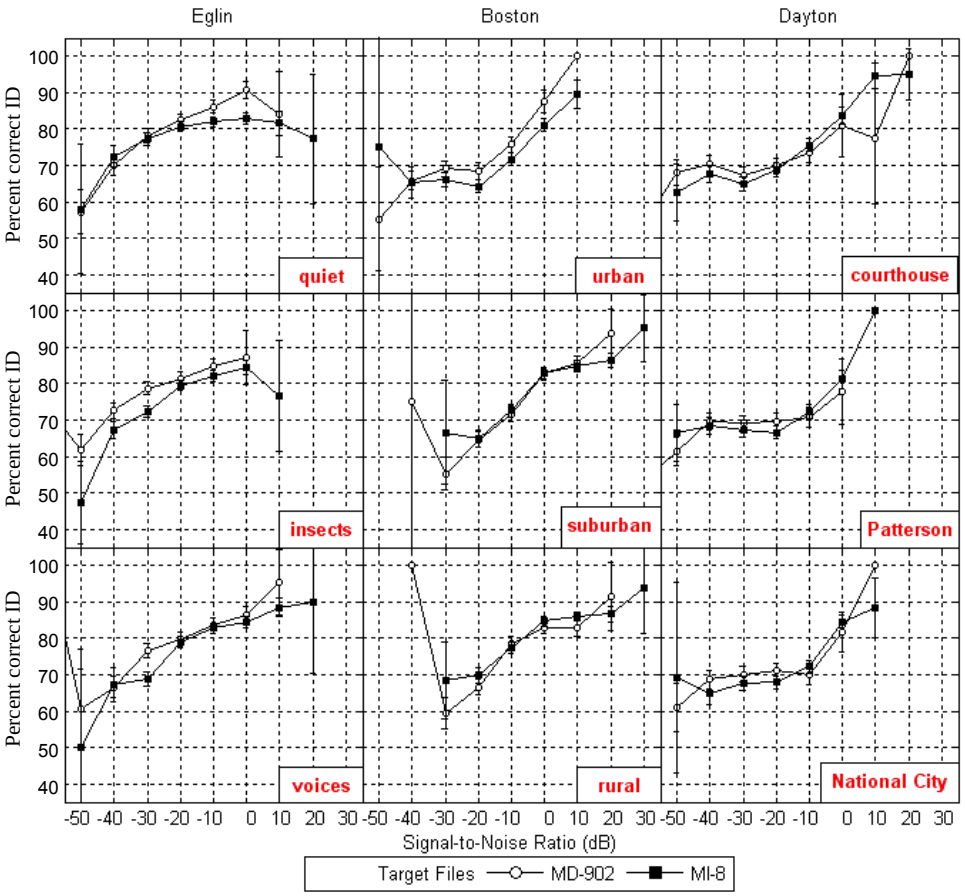


Figure 22: Classification of helicopters in different ambients.

Comparison of human and AUDIB results as a function of AUDIB Prediction

If the AUDIB model is properly predicting human performance, then the average level human performance across all trials that result in the same AUDIB prediction should exactly match that prediction. Furthermore, there should be no systematic interaction between the predicted level of performance, the actual level of human performance, and the type of ambient soundscape.

Figure 23 shows human performance for each ambient soundscape as a function of predicted AUDIB performance. These results were obtained by 1) calculating an AUDIB prediction for every target-masker combination that was presented to at least one listener in the experiment; 2) binning together all trials that had approximately the same predicted level of AUDIB performance (in bins that were 10% wide); and 3) taking the average percent correct across all

listeners for that condition. Accurate prediction from the model should result in plots with a slope of 1 (that is, from the bottom left corner to the top right corner of the graph), and with no separation between the lines. The separation between the lines demonstrates the differences in the probability of detection as predicted by AUDIB and the human results for each of the ambient soundscapes. These results show reasonable agreement between the AUDIB Model and human performance for some of the soundscapes (in particular the three Boston soundscapes and the Dayton Courthouse soundscape). However, the AUDIB model severely overpredicted performance in the Dayton Patterson and National City soundscapes, and it underpredicted it for the Eglin soundscapes.

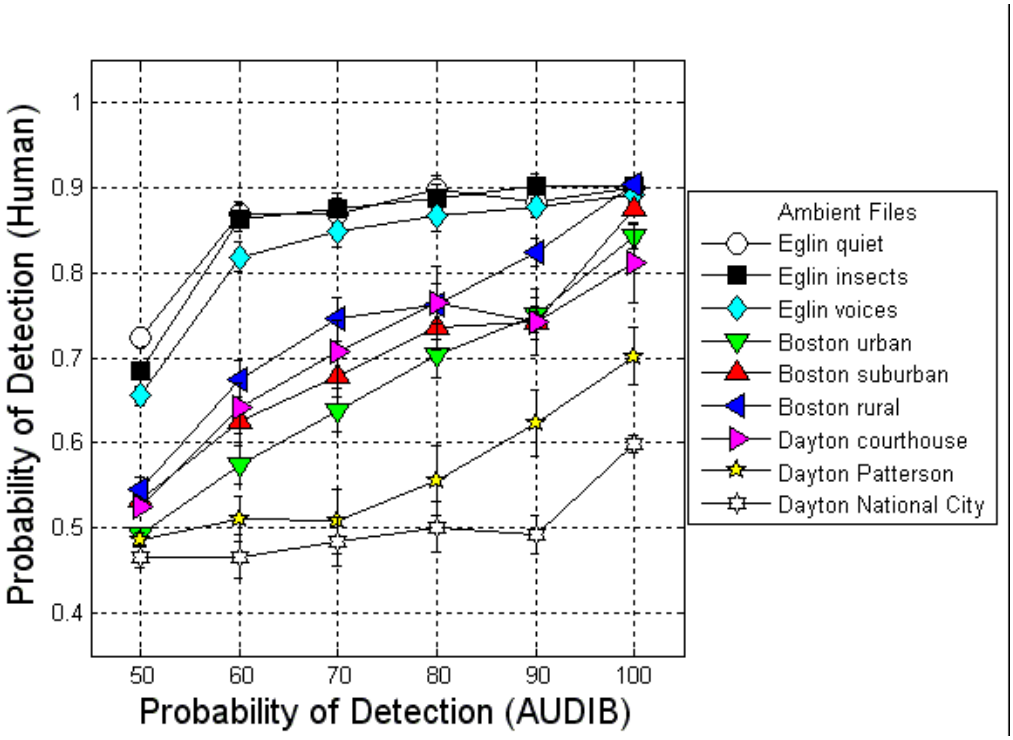


Figure 23: Correlation of AUDIB with human data (long term integration).

A short-term version of AUDIB

One weakness of the baseline AUDIB model is that it only a single, overall long-term ambient spectrum to make its audibility calculations. Thus, it does not account for short-term fluctuations in level that might make a target signal detectable in a "gap" in the masker waveform. In order to examine the extent to which this issue could explain the poor performance obtained with the AUDIB model, a modified AUDIB model was constructed that calculated the probability of detection on the basis of the masker waveform present in the 1-s interval

where the target was presented, rather than the long term ambient spectrum of the masker.

As shown for the long term ambient soundscape analysis, the correlations for AUDIB and human data are plotted with the short term integration window for each ambient soundscape. These are shown in Figure 24. This plot shows an improvement in the correlation between the human data and the model for the National City Bank ambient, however, the prediction remains limited across the ambient soundscapes overall. As for the prior correlation, a good prediction would be revealed in a slope of 1 for the plots, with no separation between them.

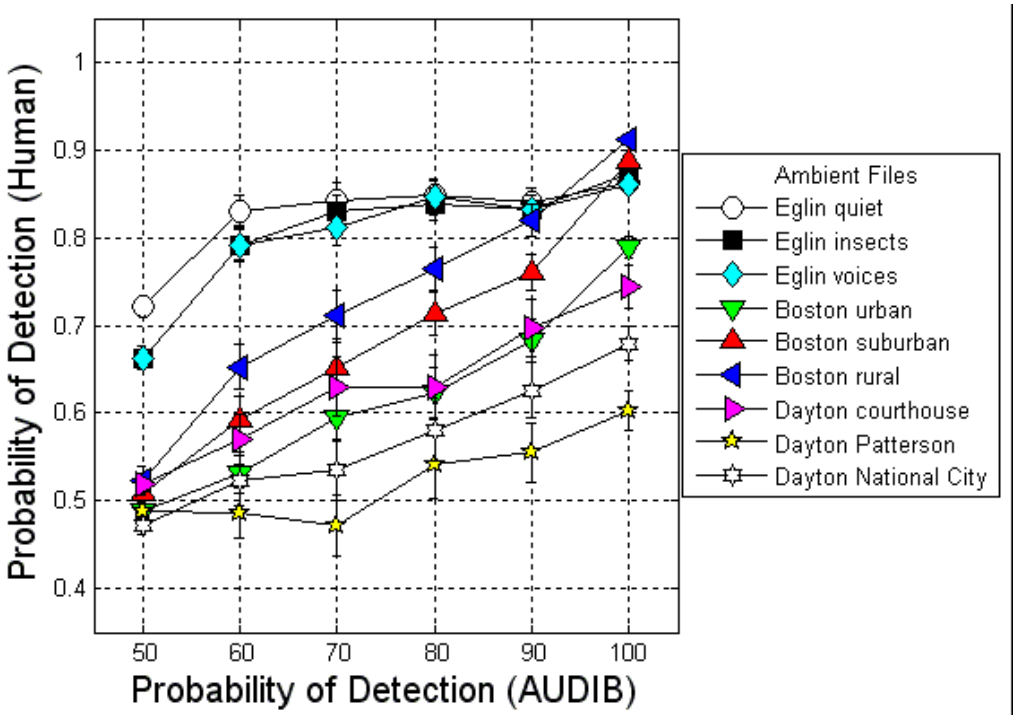


Figure 24: Correlation of AUDIB with human data (short term integration).

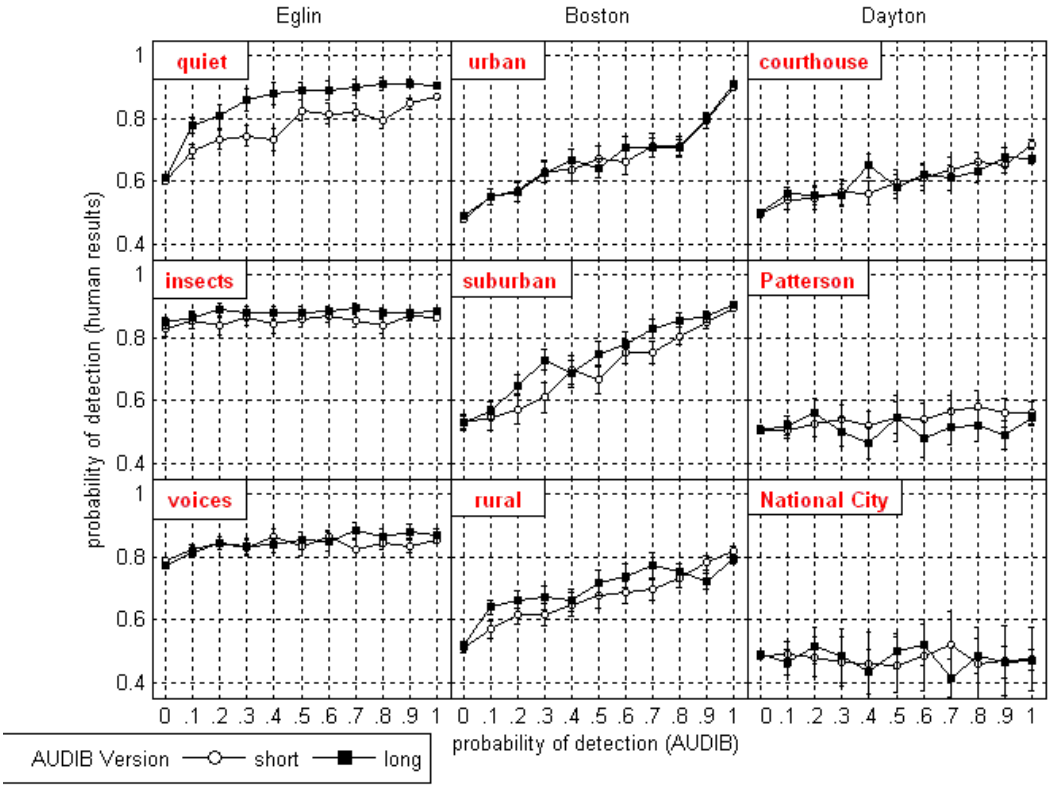


Figure 25: Comparison of AUDIB predictions for current version (long) and modified version (short).

Conclusions

The current version of the AUDIB model for detection of rotorcraft has been shown to have significant limitations. The predictions based on this model appear to be a reasonable match to the human listener results for some ambient soundscapes, but not all. A problem that was found in the implantation of the model was a limit in the sensitivity of the filters for the low frequency bins. The model bases the processing on linear frequency bins, rather than logarithmic frequency, causing it to exhibit poor sensitivity to differences in the low frequencies. An additional difficulty that was found in the procedures used here was the use of flat weighting for calculation of the signal-to-noise ratios. While the low frequency components affected by this are inaudible to humans, and presumably to the model, by excluding the frequency range below approximately 20 Hz, the model can offer better predictions for the low frequency signals. This could be accomplished simply with a high pass filter integrated for this range. Further improvement of the model for a variety of ambient background settings would account for the spectral and temporal differences in environments, rather

than simply the intensity levels. As shown in this study, the spectral components of the background, as well as the temporal variation, provide challenges to the performance of the model. The urban settings, with their motor noises and reverberation, provided a significant problem for the predictive ability of the model.

Subjects were asked to identify which helicopter was presented in the trial, and these data are plotted individually by helicopter, showing a similar accuracy for both aircraft in most SNR and ambient soundscape conditions. This basic level of classification indicates that neither helicopter was more easily identified, and thus classification was not systematic for these signals in these conditions.

Further examination could include more specific comparison on the basis of the different spectra in the ambient sounds and the target signals for each experimental interval. This analysis could provide information about what components in the noise most influence detection of the targets. Further, comparison of variability in the amplitude modulation of different ambient soundscapes should reveal information related to specific detection thresholds. These analyses are possible based on the data from this study. This analysis would also be extended to the classification of the target signals.

Another extension to the current study would address localization of the target sounds within the ambient soundscape. This could not be completed at this time due to limitations in documentation related to the flight paths of the helicopters at the times the signals were extracted. Thus, the reference location could not be generated, and the accuracy of subject responses could not be determined. Plans are for this completion of this work to be done in the near future.

The data collected in this study also provides a framework for evaluation of detection models like AUDIB, and will function as a test bed for future model development.

References

Abrahamson, A.L. (1975), “Correlation of actual and analytical helicopter aural detection criteria, volume 1,” U.S. Army Air Mobility Research and Development Laboratory, Technical Report #74-102A.

Elshafei, M., Akhar, S., Ahmed M. S. (2000), “Parametric Models for Helicopter Identification Using ANN,” IEEE Transactions on Aerospace and Electronic Systems, 36(4).

Fletcher, H. (1940). “Auditory patterns,” Reviews of Modern Physics, 12, p. 47-66.

Fidell, S. (1977). “Relationship between detectability and annoyance of low-level signals,” J. Acoust. Soc. Am., 62(S24).

Fidell, S. and Bishop, D.E. (1974), “Prediction of acoustic detectability,” U.S. Army Tank Automotive Command, Technical Report #11949.

Fidell, S., Pearsons, K.S., and Bennett, R.L. (1972), “Predicting aural detectability of aircraft in noise backgrounds,” A.F. Flight Dynamics Laboratory, Technical Report #AFFDL-TR-72-17.

Fidell, S., Pearsons, K.S., and Bennett, R.L. (1974), “Prediction of aural detectability of noise signals,” Human Factors, 16(4), p. 373-383.

Green, D.M. (1959), “Auditory detection of a noise signal,” J. Acoust. Soc. Am., 32(1), p. 121-131.

Green, D.M. and Swets, J.A. (1966), Signal detection theory and psychophysics, John Wiley and Sons, Inc.

Greenwood, D.D. (1961), “Auditory masking and the critical band,” J. Acoust. Soc. Am., 33(4), p. 484-502.

Hartman, L. and Sternfeld, H. (1973), “An experiment in aural detection of helicopters,” U.S. Army Air Mobility Research and Development Laboratory, Technical Report #73-50.

Horonjeff, R., Fidell, S., Green, D. (1983), “The detectability of repetitive, periodic impulses,” U.S. Army Research Office, Technical Report #16729.5-LSI

Loewy, R. (1973) “Aural detection of helicopters in tactical situations,” Department of mechanical Engineering, University of Rochester, N.Y.

Jamieson, L.H. (2002), “Speech Processing by Computer,” Purdue Electrical and Computer Engineering Course, Purdue University, West Lafayette, IN 47907-1285.

Ollerhead, J.B. (1971), “Helicopter aural detectability,” U.S. Army Air Mobility Research and Development Laboratory, Technical Report #71-33.

Poulson, T. (2007) Ear, hearing, and speech, *Fundamentals of Acoustics and Noise Control* (Jacobsen, Poulsen, Rindel, Gade, and Ohlrich), DTU Technical University of Denmark, p. 72.

Moore, B.C.J. and Glasberg, B.R. (1983), “Suggested formulae for calculating auditory-filter bandwidths and excitation patterns,” J. Acoust. Soc. Am., 74(3), p. 750-753.

Selvy, R. (2002) “Development of the Matlab-based MICHIN helicopter aural detection model,” Master’s thesis, Naval Postgraduate School, Monterey CA.

Taylor, D. and Poe, A. (1973) “An Aural Detection Model,” U.S. Army Missile Command, Technical Report #RD-73-11.

Ungar, E.E. (1971), “A guide for predicting the aural detectability of aircraft,” A.F. Flight Dynamics Laboratory, Technical Report #AFFDL-TR-71-22.

Zwicker, E. and Fastl, H. (1972), “On the development of the critical band,” J. Acoust. Soc. Am. 52(2), 548-557.

Zwicker, R., Flottorp, G., and Stevens, S.S. (1957). "Critical bandwidth in loudness summation," J. Acoust. Soc. Am. 29, 699-702.

APPENDICES

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF EQUATIONS

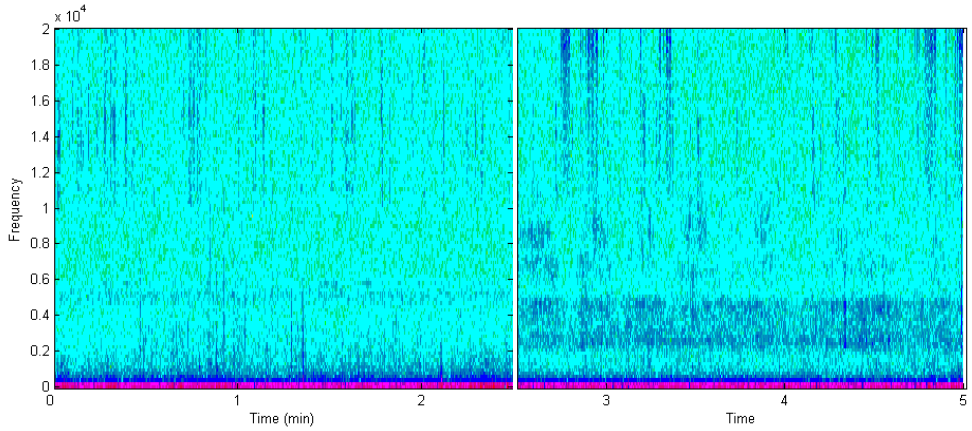
LIST OF TABLES

Appendix 1

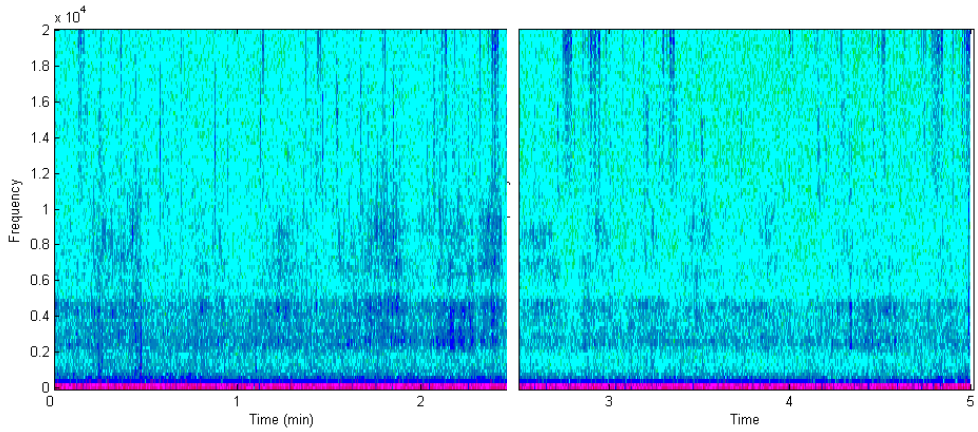
Ambient soundscapes

Table 1-1: Ambient soundscape descriptions

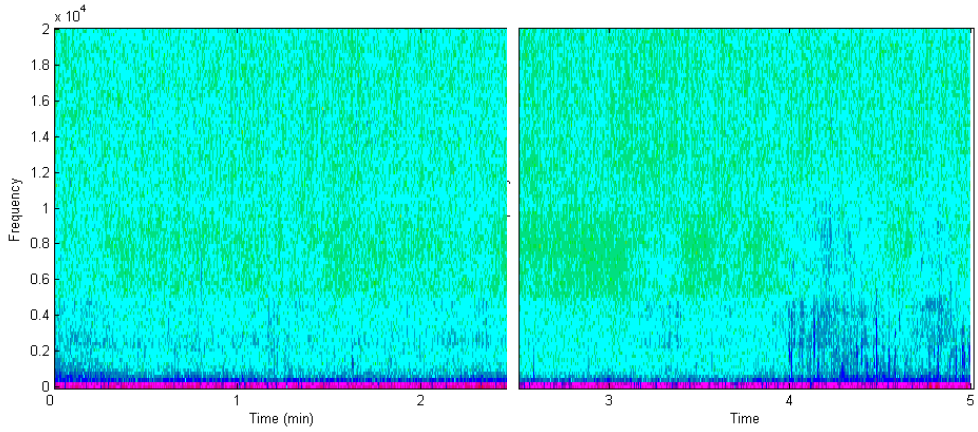
ambient soundscape	description
Ambient 19 – Eglin quiet	quiet, open field
Ambient 5 – Eglin insects	field with insects, birds
Ambient 28 – Eglin voices	field with some voices, clothing rustling
Boston urban	vehicle traffic, alarms, set back in Boston Common
Boston suburban	automobiles, pedestrians, Newton, MA
Boston rural	distant aircraft and trucks, birds, Boxford, MA
Dayton courthouse	set back from street, heavy vehicle traffic, semi-urban canyon, 3rd Street and Main (in downtown)
Dayton Patterson	3rd Street and Patterson, downtown urban at sidewalk level, heavy vehicle traffic
Dayton National City	ATM vestibule immediately outside main entrance to the bank, 3rd street in downtown, highly reverberant flat stone surfaces, heavy vehicle traffic



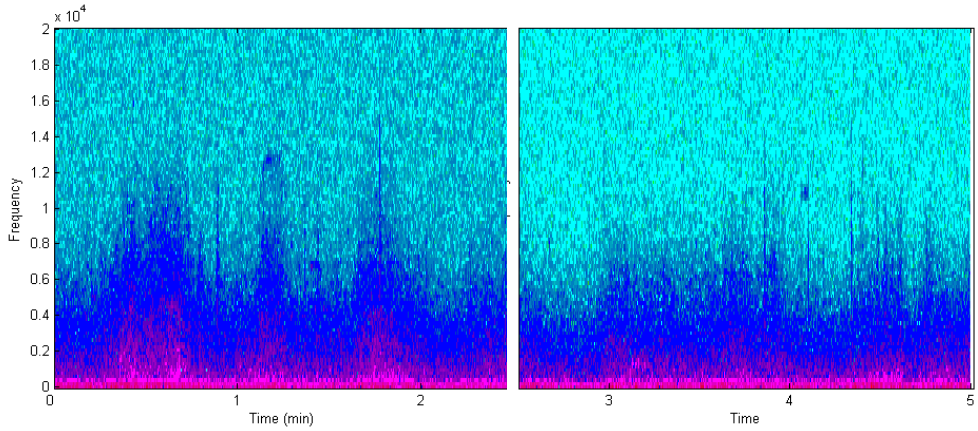
Eglin quiet



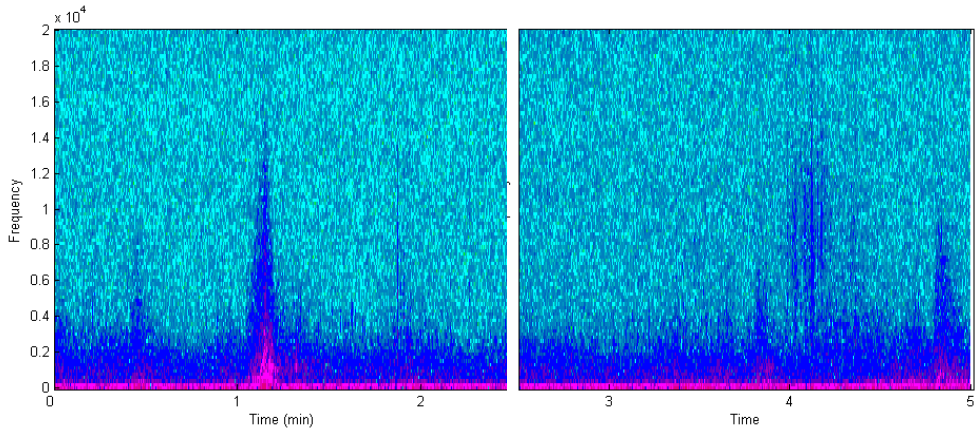
Eglin insects



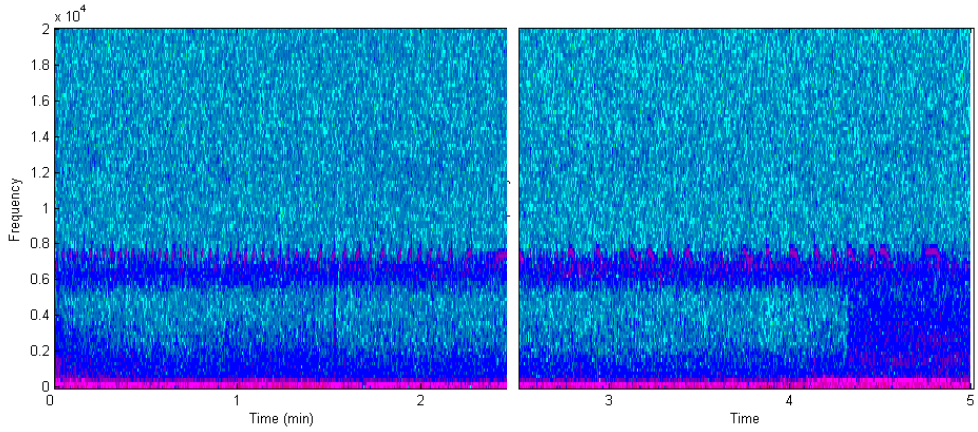
Eglin voices



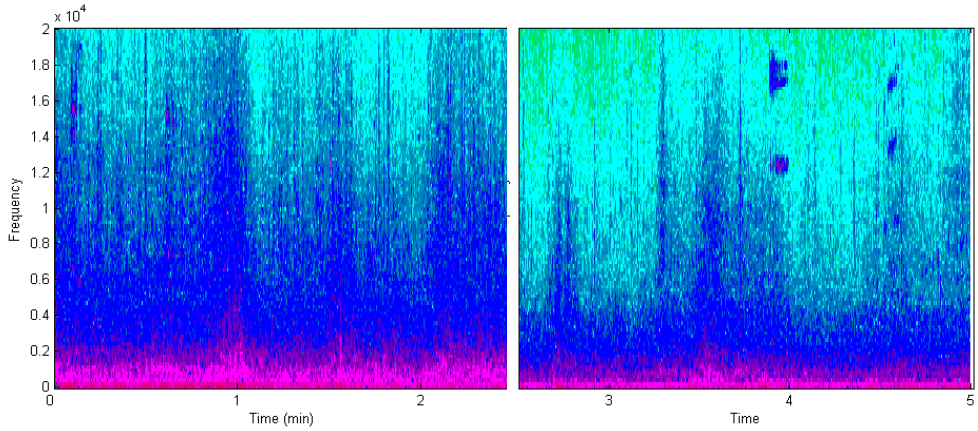
Boston urban



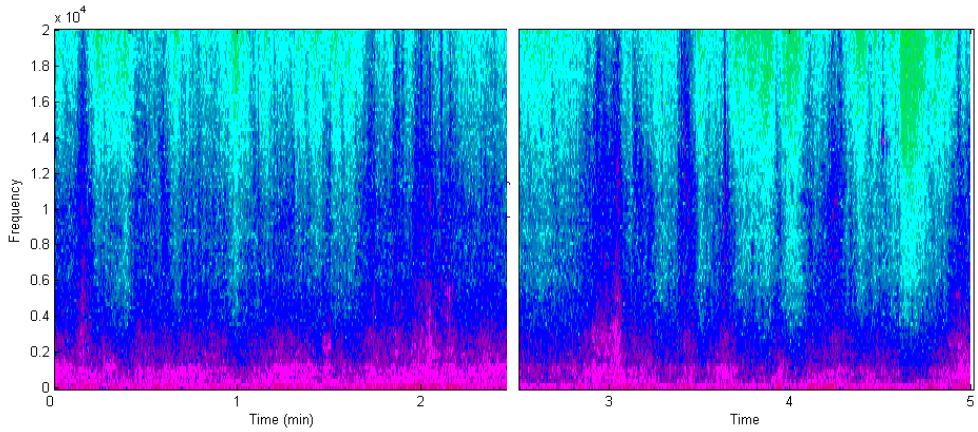
Boston suburban



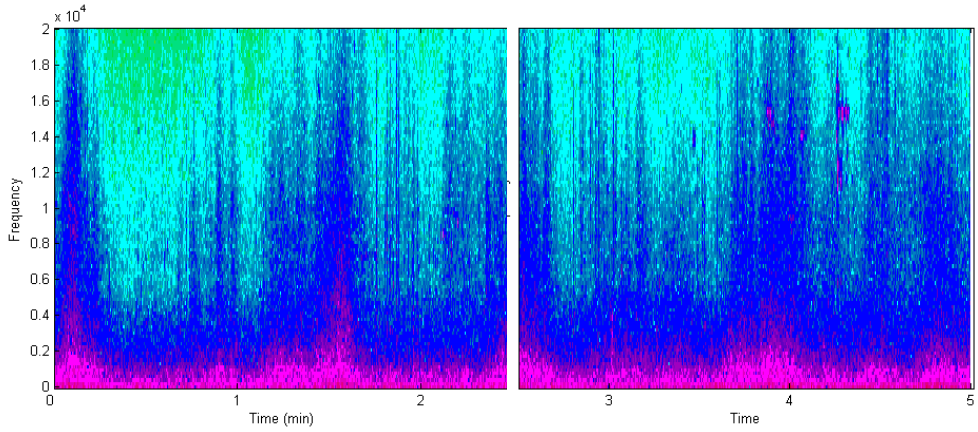
Boston rural



Dayton courthouse



Dayton Patterson



Dayton National City Bank

Figure 1-1: Spectrograms of the ambient soundscapes. Light blue and green represent the least intense spectrum, dark blue is greater intensity, fuchsia is the greatest intensity.

Appendix 2

Psychoacoustic specifications

Headphone responses

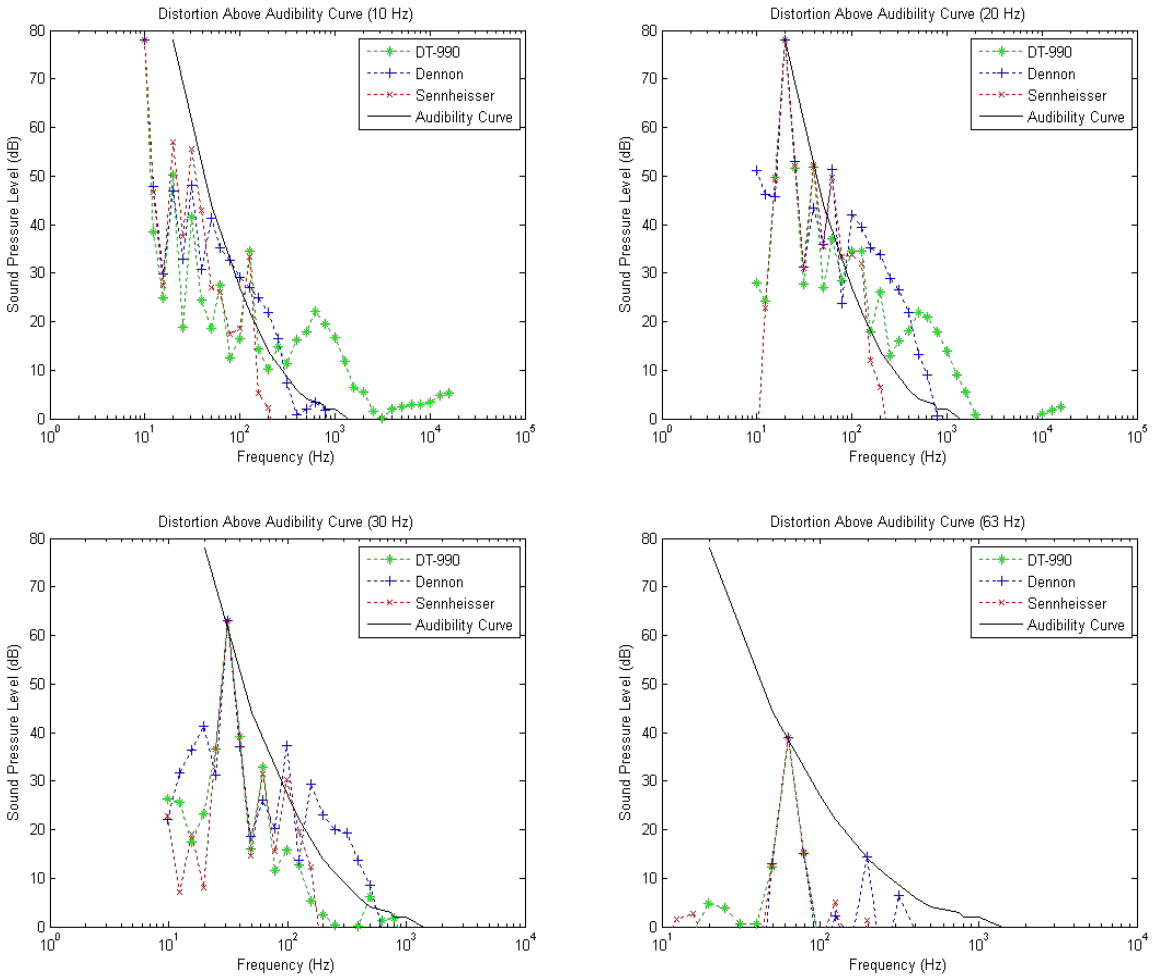


Figure 2-1: Headphone response curves for BeyerDynamic DT-990 (selected for this study), Denon AH-D1000, and Sennheiser HD280 pro headphones plotted against the human audibility curve. Each panel represents an input frequency as indicated: 10 Hz, 20 Hz, 30 Hz, and 63 Hz.

Filter weighting

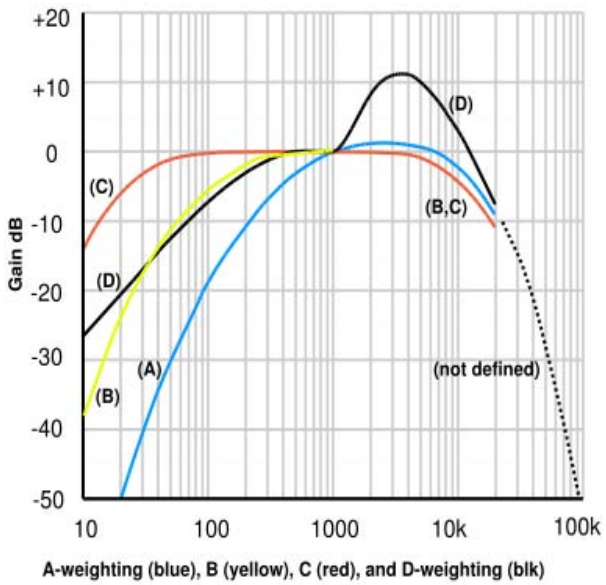


Figure 2-2: Sound level filter weighting functions

Peripheral auditory processing in humans

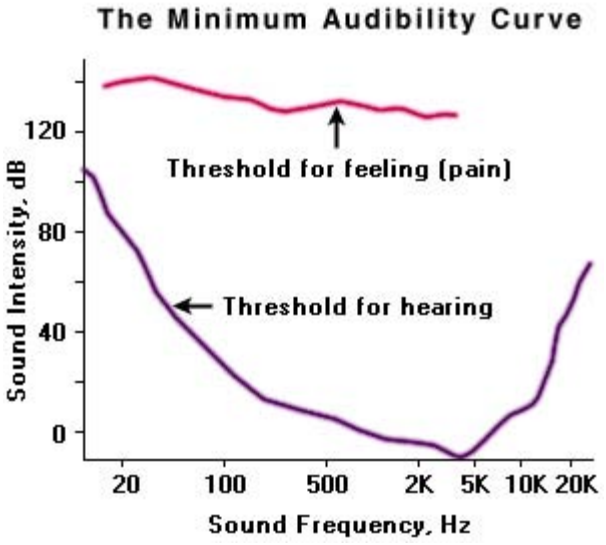


Figure 2-3: Human audibility curves

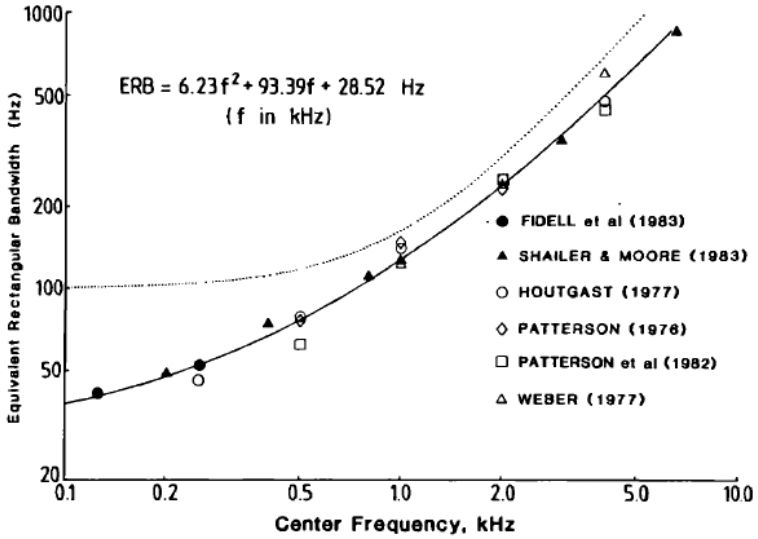


FIG. 1. The symbols indicate estimates of the equivalent rectangular bandwidth (ERB) of the auditory filter at various center frequencies, taken from the results of the workers indicated. The curve fitted to the data is specified by the equation in the figure. The dotted line is the classical critical-band function.

Figure 2-4: Equivalent rectangular bandwidths for filters in the human peripheral auditory system. Taken from Moore and Glasberg (1983).

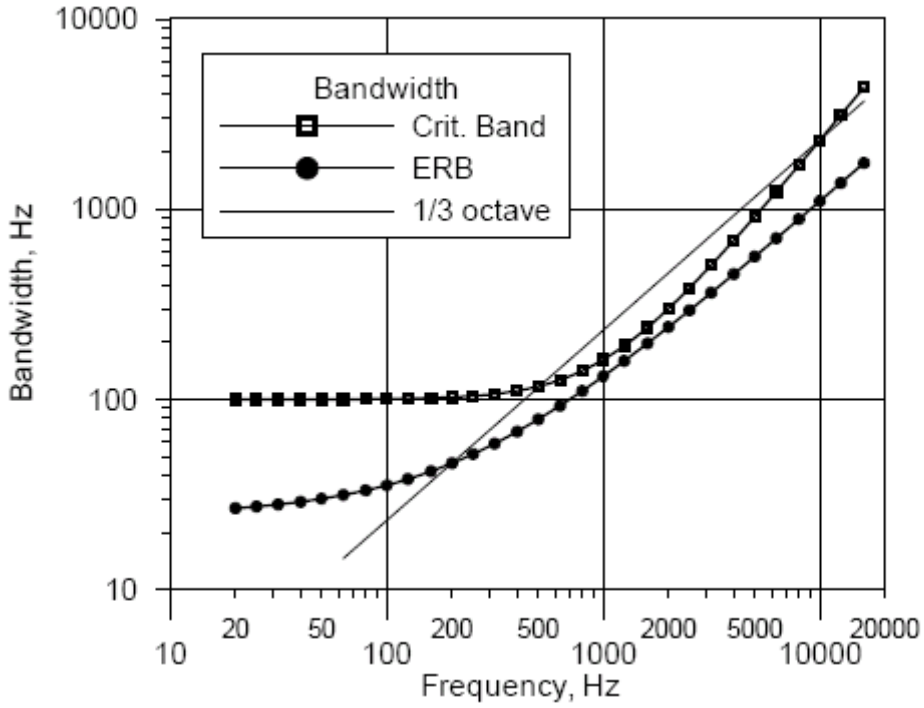


Figure 2-5: Bandwidth of critical bands and Equivalent Rectangular bandwidth, ERB. The bandwidth of 1/3-octave filters (straight line) is shown for comparison. Taken from Poulsen (2007).

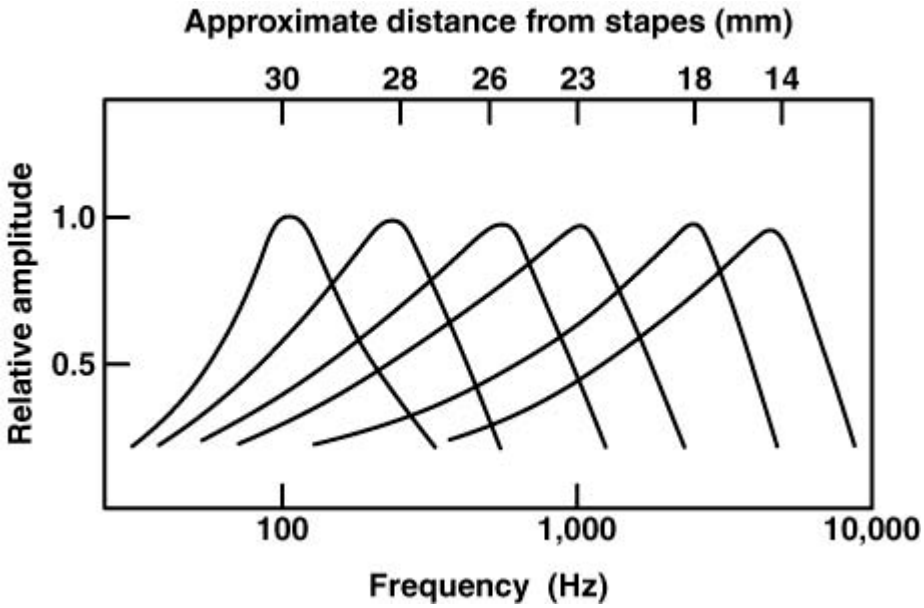


Figure 2-6: Response of the basilar membrane in the cochlea as a function of stimulus frequency. The critical band/ERB calculation is based on this response. Taken from EE649: Speech Processing by Computer website, Purdue University (2002).

Appendix 3

AUDIB model prediction

A thesis from the Naval Post Graduate School describes the implementation of the AUDIB predictive audibility code in MATLAB® (Selvy, 2002). A review of that work and AFRL's own independent analysis of the AUDIB original FORTRAN code enabled the development and execution of a new MATLAB® version of AUDIB in this experiment as a "virtual listener".

Analysis of Rotocraft Noise Model Audibility Module.

This module calculates the human audibility at a single receiver point. It was initially developed by John Ollerhead at Wyle Laboratories. In 1975 it was implemented in I Can Hear It Now (ICHIN). The current code is derived from the 1986 version of ICHIN 6 which was developed by NASA. The current model reads the time history data from an ASCII file. Though the time history may have discontinuities, the frequency spectrum must be continuous. It assumes that the background is uniform in time, but the ambient levels can be specified in the first line of the spectral data to give the user the ability to vary the ambient level by location. This implementation of human detection is based on the methodology defined in USAAMRDL-TR-74-102A by Ollerhead. The d` metric was added to the computations in 2004 by Wyle Laboratories. This is based on the US Park Service Grand Canyon project. It is implemented for the one-third fractional octave bands from 50 Hz - 10,000 Hz.

The method uses critical bands and signal-to-noise information to determine the probability of detection (POD). The receptors used are characterized by a single listener or group.

Obvious limitations of the FORTRAN implementation:

1. filename and pathnames limited to 1024 characters
2. input narrow band spectra limited to 2048 frequencies

A call diagram is shown in Figure 3-1.

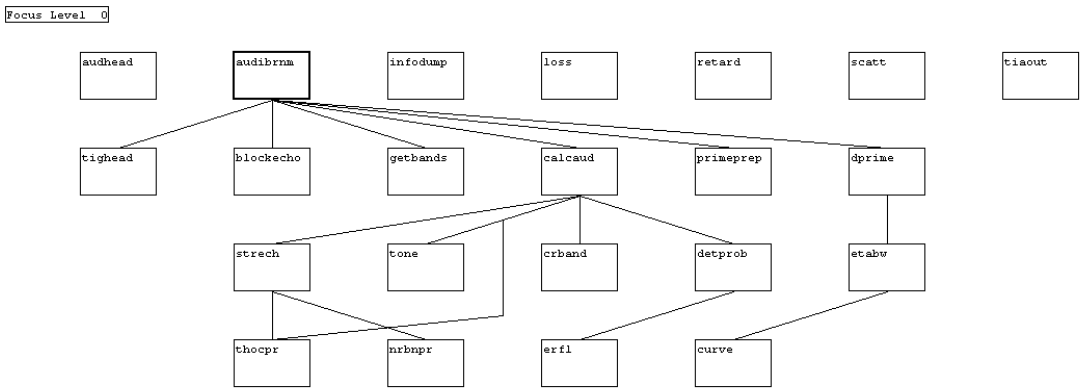


Figure 3-1: Call tree for AUDIB_RNM.

Most of the functions that are listed in the first row have no lines from them, as they are not called when executing the audibrnm program. We will examine these functions first, since they can be removed without impacting the overall function of the detection program, audibrnm.

AUDhead

This function writes the header of an already opened TIA file. The TIA file is an output of AUDIBRNM. This is used in an additional model to determine hot spots call SPAR.

Infodump

This function dumps the input into an output file.

Loss

This function calculates the loss in sound pressure level due to atmospheric absorption. This is no longer needed since this is done through RNM.

Retard

Here measured slant range, altitude and velocity are converted using time-retarded coordinates. This is most likely also a holdover from the ICHIN program that did some of the propagation that RNM is now responsible to provide the audibrnm program.

Scat

This is another propagation function. It computes the atmospheric absorption based on inhomogeneities.

TIAout

This dumps the detection information to the opened TIA file that was created as part of the TIAhead function.

Audibrnm

Now that the analysis of the extra functions is complete we can examine the flow of the audibrnm function. This is the heart of the program. The program first opens the file that was passed as a command line argument. From this file the name of the stimulus (*.TIG) is obtained. Additionally the frequency range over which the detection will be calculated is read from the file. This is adjusted to ensure that the first and last frequencies are integer multiples of the bandwidth, which is also read from the input file. Lastly the program reads the path of the ambient data file. AUDIBRNM starts the data calculations with a **TIGhead**. This function will read the header of the propagated data at the various grid points. At this point the data is stored in a string array for processing in the **BlockEcho** function. The **BlockEcho** function reads the header for the first data section. From this header the **BlockEcho** function returns the number of time spectra in the data section and (x, y, z) position of the point. Following the **BlockEcho** is a call to **GetBands**. This function is a multipurpose function reading the inputs from multiple files and multiple types of lines. Each of these is selected by specifying the mode which **GetBands** is to operate for this call. The modes are listed below:

1. Scan the band number and get the min/max band number
2. Load the Ambient numbers for this grid point
3. Load the time and SPL spectrum
4. Populate the frequency array
5. Load the background data from the ambient file
6. Read a line from a file and do nothing with it.

The first call to the function is to get the minimum and maximum values for the frequencies in the data file. Next audibrnm writes the headers for the output files, i.e. the maxPOD, allPOD and maxDprime files.

GetBands is called again to get the ambient data line for the specific point, however, this data is not used for the analysis. If the Uniform keyword, in the ambient data, is set to a value of 1 then the program will read the ambient data for each of the ambient spectra. The next step is to read the uniform ambient from the ambient file. This is done within the audibrnm function. The file format is:

```
Comment
Uniform keyword with associated value
Number of ambient frequencies
The ambient frequency list
The sound pressure level list
```

The formats of the lists in this case are somewhat arbitrary. The example files have 10 entries per line. However, the implementation of these values is not specific on how the data should be formatted. In fact, the creation of the ambient files listing all the frequencies on one line with the associated SPL levels on the next line is equally valid, as is listing all the frequencies on separate lines with the SPL values following. After the frequencies are read, the program checks to ensure that the frequencies are appropriate based on the previous information. That is, if the calculation is to be done with one-third octave bands, the frequency list must contain 31 bands. Additionally the first band is checked to be 10 Hz. This means that the input spectrum is to be from 10 – 10,000 Hz. If the mode of calculation is narrow band the program ensures that the first frequency is equal to the frequency increment from the input file. Also the last frequency must be equal to the frequency spacing times the number of frequencies.

If the loop has been completed once already, the program skips the above step and reads the next ambient data line from the file. Next, the program calls **GetBands** to read the time and SPL history from the file. The number of time increments was specified when the program read the header of the data section. The program then reads this length of data from the file. The frequencies are checked for consistency with the specified bandwidth. If the frequency bandwidth is more than 5 Hz from the specified bandwidth then the flag is set to stretch the data. The algorithm defined by Ollerhead is completed in the **CalcAud** function. The discussion of that portion of the program follows.

The probability from the **CalcAud** function is written to the maxPOD and all POD file. If the data is fractional octave with the number of frequencies equal to 31 the function **PrimePrep** is called on the data. All this does is to eliminate the elements of the 10 - 10,000 Hz array that are without the 50 – 10,000 Hz range of the d` calculation. Unless there was an error in the **PrimePrep** the **Dprime** function is called. The results of this function are written to the output file. Once this is completed the next data line is read.

CalcAud

This function is to calculate the audibility of the SPL spectrum against the ambient. It is based on the implementation of a method described by Ollerhead in 1974. The calculation procedure as described in the documentation is:

1. Convert the input sound pressure level spectrum to the users working power spectral density spectrum
    a. The number of requested bands is limited to 2048
    b. Frequency bandwidth cannot be less than 15 Hz
    c. Maximum frequency is 8,000 Hz
2. Initialize the Listener Criteria Critical Band differentials for detection distances (maximum, median, minimum)

    a. Based on criteria specified in the original 1974 TR (page 109)
    b. Alone(-3.0, 0.0, +3.0)
    c. Crowd (-4.0, -3.0, -2.0)
3. Compute the absolute threshold for each frequency
    a. Based on 6th order polynomial fit defined in 1974 TR
4. Compute Critical Band level for the ambient and stimulus data.
    a. Based on Greenwood's published relation for critical bandwidth
    b. Units of BARK
    c. Computed at the user specified frequencies
5. Determine audibility by summing the audibility in each band
    a. Determined by comparison of signal with the combination of the background and absolute tone threshold.
6. Use listener criteria to assign a probability of detection

For the start of the procedure, if the resolution is fractional octave and all of the stimuli are below the level of the background the function returns zeros. Next, the levels are converted from centibels to decibels. The program permits the user to provide information in different frequency resolutions in the ambient and stimuli definitions. Neither is required to have the specified frequency resolution in the input file. Regardless of whether the data is sparsely populated or not the **Stretch** function converts the narrow band information into the Master Data Format (MDF). For the fractional octave bands, this is accomplished in the **THOCPR** function. In this case the **Stretch** function is nothing more than a dead function that sends the data to be converted to the **THOCPR** function, as the number of input data is 31. This function is never called when the number of bands is equal to 31 so this code will NEVER execute. Rather the **stretch** function will only be used for the call to the **NRBNPR** function to define the MDF for the narrow band input data. It would be suggested that the **stretch** function be replaced with direct calls to the **NRBNPR** function.

After converting the data to the MDF the absolute tone thresholds are calculated across the MDF bands through the **TONE** function. This is where the 6th order curve fit is used to compute the tone threshold. Next the critical bands for both the single and crowd are calculated in **CRBand**. The audibility of the signal is determined for each of the critical bands using Equation 1.

Equation 3-1: Audibility Equation.

$$audibility_i = sig_{cr_i} - 10 \cdot \log_{10}\left(10^{amb_{cr_i}/10} + 10^{tone_{cr_i}/10}\right)$$

The probability of detection is determined by comparing the calculated audibility to the audibility of the crowd and single. This difference is sent to **detProb** where the aural detectability parameter is computed for the difference. Finally the maximum audibility is determined.

Description of the Stretch function as implemented in AUDIBRNM

The **STRETCH** function that is called at the onset of the audibility calculation is used to expand or contract the narrow band and constant bandwidth input data before the critical band calculation. **STRETCH** calls two different functions, one for the narrow band and the other for the one-third octave bandwidth. These functions are similar but will be evaluated the same way.

THOCPR – The stretching function for one-third octave band

This function as initially coded for I Can Hear It Now (ICHIN) in 1986. It has been updated twice since its initial inclusion. It is meant to expand or contract a one-third octave spectrum.

First we define a value called factor. This is $2^{1/6}$ and represents the distance from the center frequency to the upper cut-off frequency defined in the ANSI standard for fractional octave bands. The upper frequency is then determined by multiplying the center frequency by the factor. Rather than determining the bandwidth by computing the lower frequency, i.e. an additional factor of $2^{-1/6}$, the bandwidth is determined by finding the difference in two adjacent upper frequencies. Finally, the upper frequency is copied to a variable call FMOST.

We assign the value of FREQ to be an integer multiple of the desired resolution, RES2 prior to starting through the calculation loop.

For j = 1:length of new array
Desired frequency is copied from the array
Lower limit of that frequency is determined by subtracting half of the desired resolution from the desired frequency.
Upper limit is determined by adding half of the resolution
If (the upper limit is greater that the last given frequency)
Make the value at j equal to the value before it…

Determine the index of the fractional octave band the lower limit falls into
Determine the index of the fractional octave band the upper limit falls into
If (ilow and iupper are equal)
Level at j is equal to the level at ilow/iupper minus the bandwidth of the ilow/iupper band
Else
Determine the fractional part of the energy that is above the lower boundary (EL1)
Determine the fractional part of the energy that is above the upper boundary (EL2)
Sum the energy in the bands between the ilow and iupper. (EL3)

The j^{th} element is $10*\log_{10}(EL1 + EL2 + EL3) – NewBandwidth$

After the examination of the ThoCPR function it was determined that the differences between the ThoCPR and NrbNpr functions are when the bandwidth is incorporated in the calculations. Otherwise the flow of the program is the same.

Appendix 4

Hidden Markov Modeling:

There have been attempts to build parametric artificial neural network classification models for helicopter identification and classification in the technical literature (Elshafei, Akhtar, and Ahmed, 2000).

For this experiment, preliminary development of a Hidden Markov Model (HMM) based classifier was undertaken to investigate the potential for predictive classification for type of helicopter as well as other noise sources based solely on auditory signals. Preliminary results indicate that this HMM based method is very attractive for classification of helicopter audio signals. Using the signals of two helicopters, i.e. the MD-902 and the MI-8 for training, the HMM based classifier was able to correctly classify 95% of the test data files.

Statistical methods are a good technique to classify such processes as they have very good recognition ratios. One statistical method that is used extensive is the Hidden Markov Model (HMM).

HMMs are used for source classification because of the positive results in recognition ratios, based on the statistical method employed. There are barriers to their use, however, including the large numbers of computations required and, to a lesser extent, the large amount of memory required. These make implementation on notebook PCs problematic, where the computing and storage resources are constrained. When the unit operates in a low signal to noise ratio environments, lower classification ratios also become a significant problem.

The overall view of the HMM is given in Figure 1. This has been organized into three stages. The input enters the model at the first stage (*Stage A*). This stage has been trained to classify an assortment of signals, including trucks, helicopters, motorcycles, etc. Once an input stream is classified as a helicopter signal, it is used as input for *Stage B1.* In this stage the input stream is classified according to type of helicopter (in this study, MD-902 or MI-8). In *Stage C1* and *Stage C2* the input is classified as moving towards the observer or moving away from the observer. Note that *stages B* and *C* have only the HMM in them. They do not require linear predictive coding (LPC) and vector quantization (VQ) (described in the following section) and can use the same code book.

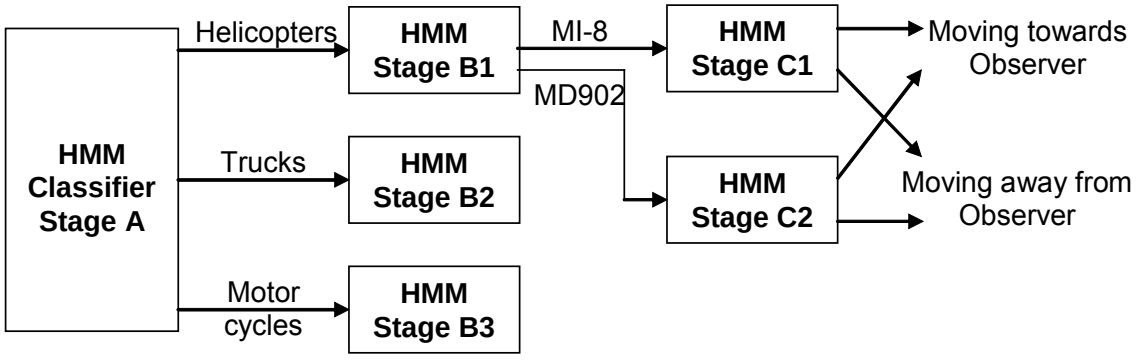


Figure 4-1: Major blocks of the proposed classification tool

Basically, the input stream is classified in multiple HMM stages that are trained to various levels of information that is required. This design reduces the computations that are required and also increases the classification ratio. It also increases the throughput of the system to classify a given input stream by hosting each of the stages on a single core of a processor.

The major blocks of HMM *stage A* are given in Figure 2. The signals are considered as input through a microphone and converted to digital values using an A/D converter (not shown in figure). The digital signals are processed through a series of noise filters (High-Pass Filter and Low-Pass filter) to attenuate the noise. It is passed through an LPC module and a VQ module. The resulting codes are input to the HMM block (*stage B*) that determines the classification of the input within the larger category.

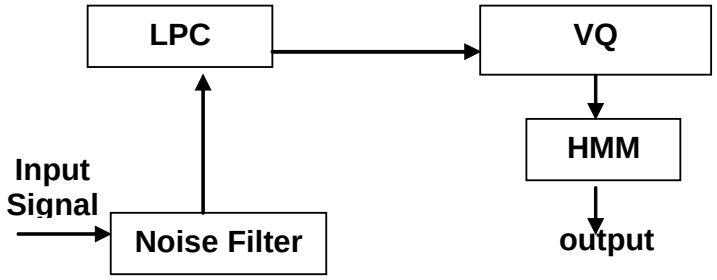


Figure 4-2: Block diagram of the major components of HMM classifier Stage A

Linear Predictive Coding (LPC)

LPC is used to analyze signals by estimating the spectral peaks, removing the effects of the peaks from the input signal by inverse filtering and estimating the intensity (or power) and frequency (pitch) of the remaining signal, to arrive at the residue signal. A difference equation is used to determine the peaks from the signal by expressing each sample of a digitized input signal as a linear combination of representations of previous samples. The coefficients of this difference equation characterize the peaks, which are estimated. For example, the maximum coefficient can be selected as the spectral peak for a given signal.

LPC is used to remove the distortion in the input signal caused by variations in signal quality. The underlying theory behind the LPC model is that a given sample can be approximated as a linear combination of past samples. Mathematically the value *s(n)* of a sample at time *n* could be given as:

Equation 4-2: LPC equation

$s(n) \cong a_1 s(n-1) + a_2 s(n-2) + a_3 s(n-3) + \ldots + a_p s(n-p)$ using previous samples.

For this study, 10 previous consecutive samples were used to predict the sample value. The main objective was to determine the coefficients, a_k, which are the linear predictor coefficients that give the minimum error to represent an input window of 180 samples.

The basic problem is to determine the set of linear predictor coefficients from the digital input signal within an analysis window of 180 samples. These coefficients are determined so that they match the properties of a digital filter. Since the spectral characteristics of the input vary with time, the predictor coefficients must be estimated from a short interval of the signal (180 samples) around a given time. The approach is to find a set of predictor coefficients that minimize the mean-square prediction error. The standard methods to determine the prediction coefficients are the autocorrelation method and the covariance method. In addition to 10 LPC coefficients, we use the power and the pitch of the 180 samples of the analysis window to characterize the samples. These 12 coefficients are considered as the feature vector and used in latter stages.

Vector Quantization (VQ)

The output of the LPC analysis is a series of feature vectors, which consist of the predictor coefficients of the LPC analysis and the root mean square energy of the input (the power), and pitch, characterizing the signal of a given window.

The vector representation after LPC reduces the required number of samples from all possible combinations of coefficients to a 12-tuple-feature vector that represents the 180 samples of the analysis window. A single representation for each input is ideal, since each input contains more than one 180-sample window. A codebook of vectors is generated from these representations with significantly more code items. This is commonly called *vector quantization*. Each feature vector is represented by a discrete symbol, determined by the codebook of vectors that were generated by the training set. The codebook vectors represent a given set of signals used in the training, which are subsequently used by the classifier. During training, all the input signals must represent as many targeted scenarios as possible and codes are developed based on the input. These codes are stored in a codebook to be used by the classifier to be as complete as possible for classifying auditory signals.

Hidden Markov Model (HMM)

One of the most attractive statistical methods and frequently used techniques for classification is the HMM approach. The underlying assumption of the HMM is that the input signal can be well characterized as a parametric random process, and the parameters of the stochastic process can be determined in a precise, well defined manner. This has been shown to be a highly reliable way of classifying in a wide range of applications.

Basically, in a Markov model each state corresponds to a deterministically observable event. Thus, the output produced by the sources in any given state is not random. Since this is very restrictive, this concept is extended to include the case in which the observation is a probabilistic function of the state. The resulting model is a doubly embedded stochastic process with an underlying process that is not directly observable, but can be observed through another set of stochastic processes that produce the sequence of observations.

Mathematically, an HMM can be characterized by the following:

1. *N*, the number of states in the model. Generally the states are interconnected in such a way that any state can be reached from other states.
2. *M,* the number of distinct observation symbols per state, i.e., the total number of code items, (512 for example).
3. The state transition probability distribution, A.
4. The observation symbol probability distribution, B.
5. The initial state distribution, π.

Thus, a complete specification of an HMM requires specification of two model parameters, *N* and *M*; specification of observation symbols; and, the specification of the three sets of probability measures A, B, π.

Experiments Conducted

We have conducted experiments to determine the efficacy of the proposed tool. Existing audio data of the MD-902 and the MI-8 helicopters were used as input data to train the HMM and the test. The data are the same recordings used for the human detection study. This data was segmented to one second audio files and converted from floating point wav-file format to unsigned 8-bit integer raw format. Fifty percent of the data files were used for training, and the rest was used for testing. The selection of the training data files was varied and the HMM classifier was trained and tested in each case using the specified set of data files. The classification ratios obtained in each of the experiments are given in the following table. The HMMs were customized for each of these cases. They are not the same settings for each of the experiments. The parameters that were customized are the number of state and the number of VQ code items.

Table 4-1: Classification ratios achieved in experiments conducted

	Training	**Testing**	**Correct classification ratio**
Experiment 1	First 50 %	Second 50 %	88%
Experiment 2	Odd 50 %	Even 50 %	95%
Experiment 3	Extreme 50 %	Middle 50 %	95%
Experiment 4	Middle 50 %	Extreme 50 %	91 %

Each experiment took approximately 40 minutes for training and testing. Each one second data file require less than 200 msec to be classified. These latencies are based on an AMD Athlon 4800 based machine running at 2.4 GHz, using a single core of a dual core processor.

Results of the experiments show a good accuracy for the classification of these two helicopters based on a limited number of samples, regardless of how the samples were organized for training and testing. Given additional sample sounds, it is anticipated that the accuracy for the model classification would increase further.

Appendix 5

Abbreviations/Definitions

2AFC	Two Alternative Forced Choice
AFRL	Air Force Research Laboratory
ANSI	American National Standards Institute
d prime (d')	Detection metric, independent of bias
DARPA	Defense Advanced Projects Research Agency
dB	Decibel
FFT	fast Fourier transform
FORTRAN	Computer language
HL	Hearing level
HMM	Hidden Markov Model
HRTF	Head Related Transfer Function
Hz	Hertz
ICHIN	I Can Hear It Now (computational model)
KEMAR	Knowles Electronic Mannequin for Acoustic Research
LPC	Linear predictive coding
MAF	Minimal Audible Field
MAP	Minimal Audible Pressure
MATLAB	Computer language
NASA	National Aeronautics and Space Administration
NI	National Instruments
POD	Probability of Detection
RMS	Root-mean-square
RNM	Rotorcraft Noise Model (computational model)
SNR	Signal to noise ratio
SPL	Sound Pressure Level
USAAMRDL	U.S. Army Aeromedical Research & Development Laboratory
VC	Vector Quantization
VI	Visual Interface