Abstract

This project will attempt to identify jet aircraft noise from a variety of environmental noise sources. In recent years, real estate values have become affected by their proximity to airports. With the increase in air travel, aircraft noise during takeoffs, landings, and fly overs have become a common event in many parts of the country. Simple noise level monitoring has been used to create noise maps in areas around airports, however the monitoring technique is unable to distinguish between various environmental noise sources. Typically, aircraft noise is many seconds in length, which will require the classification technique to understand the long temporal structure of aircraft noise in addition to the short term spectral characteristics. Hidden Markov Models (HMM) will be used in this project to provide the classification system a method to account for the rise, level, and tail off of the spectral events during an aircraft fly over.

Related Research

Extensive research into noise recognition has begun recently. Automatic Noise Recognition (ANR) systems have been implemented using statistical methods in order to identify various environmental sounds. Typical industry applications for these system would include: personal PDA devices, noise monitoring, animal tracking, hearing impaired assistance. Christophe Couvreur [4,5,7] is a leading researcher into environmental noise detection. His systems have employed both statistical methods, and HMMs to identify a wide variety of environmental sounds. A new breed of "intelligent" noise monitoring solutions [5] are being developed that will automatically adapt the recognition system to the environment it is placed into, and enhance the models based on the new data.

Introduction

In order to understand the approach needed to classify aircraft noise, it is important to first identify characteristics of the noise spectrum. Shown below in Figure 1 and Figure 2 are the spectrograms of two different jet aircraft fly over sounds.

![Figure 1](image1.png)

![Figure 2](image2.png)
As can be seen from the above spectrograms, the pitch of the aircraft sound changes as the plane travels overhead. Much of this change is due to the Doppler effect, which will change the frequency of the sound waves when the airplane begins to get closer to the microphone, and then continues to move away from the microphone. Interestingly however, the frequency of the strong track is different between the two noise recordings. The first recording has the pronounced track starting at 3300 Hz, while the second recording has the pronounced track at 3700 Hz. I believe that this is due to two effects. First, it is not clear as to the actual model of the plane that created these sounds. It is likely that two different airplanes are represented in the above spectrograms, and that their engines are different. Different engines are likely to have a different sound spectrum due to the variations in rotor design, and airflow patterns through the engine.

The second difference is that the angle of fly over, or speed of the aircraft with respect to the microphone location can cause a frequency shift. The second plane has a higher pronounced frequency to begin with, which could represent that this plane flew directly over the microphone location, which via the Doppler effect, would push the sound waves closer together, thus providing a higher frequency track. It may have also been traveling at a faster rate of speed. The first plane might have been 1/4 mile away from the microphone source at it's closest point, so as it approached, the sound waves were not as tightly coupled together as the second plane.

As can be seen, a classification system that will detect aircraft noise will need to be able to handle this variation in frequency tracks. It is clear however, that the frequency shifts are visible, but the time at which they happen will also vary based on the time it takes for the airplane to arrive at the closest distance to the microphone, at which time, the frequency shift should be occurring. The long temporal property of aircraft noise, makes it important to design a classifier that will account for this structure. A Hidden Markov Model will be used to provide this temporal structure to the classification system.

**Hidden Markov Models**

A Hidden Markov Model (HMM) consists of a series of states connected together to form a Markov Chain [2,3]. These states can be described by a emission probability which is typically a probability density function, and a transition probability matrix. For each state in a Hidden Markov Model, a choice can be made to transition to the same state, or to another state within the Markov Chain. The probability density function is typically a Gaussian Mixture Model with a set of n input vectors describing that state. The transition probability matrix will describe the likelihood of the next set of input vectors resulting in a match for the same state, or a match in another state. While HMMs can be fully interlinked, so that any state can reach any other state within the model, speech and noise recognition models typically have a evolving time property which would prevent "backwards" movement within the Markov Chain. Therefore the aircraft noise system will rely on a left-right HMM which is described in Figure 3.
The states of the model are represented by the circles, with their associated state labels. For each state, there is a pdf that will describe the feature vectors for that state. Although HMMs are typically used for sequence generation, audio and speech recognition systems have used HMM to predict the likelihood that an observed sequence is represented via the HMM model.

HMM sequence recognition is determined by applying a forward recursion algorithm to calculate the log likelihood of the observed sequence fitting the model. This algorithm will greatly reduce the number of calculations needed to identify the sequence match. Although speech and audio recognition systems typically use left-right HMM which can be calculated efficiently using a Viterbi Algorithm to identify the most optimal path, the forward recursion algorithm is still effective.

**Aircraft Noise Detection**

The aircraft detection system is shown in Figure 4 below. Each of the subsystems will be described and a brief discussion on why a specific subsystem was chosen is also included.

**Training**

A selection of 15 jet aircraft fly overs was used to train the HMM model. The sounds used in this project were collected from both public domain sources, and actual live recordings of aircraft fly overs. The HMM consisted of a 5 state left to right Markov Chain, with a GMM pdf for each state. The GMMs were trained on the LPC feature vectors. The first step in training was
to run the 15 aircraft sounds through an LPC feature vector extractor, to determine the LPC coefficients up to the 10th order with a window size of 1000 points. All of the test and training sounds were sampled at 8000 samples/sec, so the windows were 125 ms in length. These feature vectors, were then run through a labeling routine that would identify the sections of the vectors that would map into each of the 5 states of the HMM.

My first attempt at training the HMM did not include the labeling subsystem as I attempted to use a pure EM training method, where I would run the entire set of feature vectors into the training algorithm, and allow it to find the state separation points. The EM algorithms were provided by Kevin Murphy's Toolkit. It was clear that the toolkits available for this training would not work, as they attempt to place all feature vectors into a GMM with multiple mixture modes, and attempt to have them diverge. The problem with this approach for the aircraft noise relates back to the spectrum. The beginning and final stages of the aircraft noise spectrum have similar structures except for the very pronounced frequency track. The toolkits available would train the GMMs on the entire set and I was not able to get them to diverge appropriately in order to accurately identify 5 separate and distinct states. In addition the initialization of the GMMs was not consistence. Therefore I never obtained similar results twice.

The approach to label or mark the feature vectors would allow me to train separate GMMs on the isolated feature vector segments. This not only produced the desired result of obtaining 5 distinct GMM pdf to provide the emission probabilities for each state, it also yielded consistent GMM initialization which allowed for results to be reproduced. The segments were marked at first based on visual identification of the spectrum, and determining approximate state transition points. This method was extremely time consuming, but yielded reasonable results on a fixed set of training data. The drawback to this method would occur if more training data were used, since the visual identification of each of the new set of feature vectors would have to occur first. As will be described shortly, this method was soon implemented via a routine that calculated the segments based on the first order MFCC coefficient for log energy.

Once the GMMs were trained, the next step was to actually train the HMM to improve the accuracy of the state transition probabilities. This step did use the EM algorithm of Kevin Murphy's toolkit to formally take the feature vectors, and use them to identify the probabilities of changing states. Upon review of typical HMM training in related works and publications, it is clear that I have implemented a hybrid approach which is neither Viterbi nor EM training, but uses both methods to train the HMMs. While this method may be unorthodox, I simply was not getting the model to converge when using the pure EM method. Although EM is guaranteed to converge at a local level, it is not guaranteed at a global level. My speculation is that the sheer number of feature vectors contained in a single aircraft fly over due to it's long temporal characteristic was providing problems for the toolkits available in the public domain. While I attempted to reduce the number of feature vectors by increasing the window size during the LPC coefficient extraction, this actually yielded worse results. It appeared that by increasing the window size to a large number would hide or smooth some of the subtle changes that occur around the point where the aircraft is directly overhead. Therefore the GMMs for the intermediate states were highly specific, and were not able to represent the differences that were seen between various aircraft.

Once the EM algorithm was finished, a HMM was described completely by it's number of states, the emission probabilities for each state, in this case it was a GMM, and the transition probability matrix which represent the left-right flow of the model. It was now time to run the
test data through the model. The test data consists of 9 aircraft fly over sounds, 5 car sounds, 5 train sounds, 3 truck sounds, 3 dog barking sounds, 1 rain sound, 1 thunder sound, 1 wind sound, and 1 music sound for a total of 29 test sounds. Each of these test sounds were run through the same LPC coefficient extraction but did not get labeled into distinct segments. The forward recursive method was used on the test feature vectors to determine a log likelihood for each of the test sounds. The log likelihood were then passed on to a decision stage which identified whether or not the log likelihood represented an aircraft noise, or not. The decision criteria was selected by first running the training data through the HMM model to determine the log likelihood for data that was used to train the HMM. An average log likelihood was generated for the 15 training sounds, and the decision criteria was originally set at 80% of the average. In practice however, the resulting log likelihood for each of the test sounds, resulted in either a number close to the average of the training data, or close to -Inf. This likely occurred due to the very high probability of remaining in each of the HMM states before transition to the next state. An aircraft sound will slowly increase in energy over time as it approaches, and the slowly die away when the aircraft leaves the areas. Typically, other environmental sounds are shorter time duration such as a dog bark, or a car passing, or they are extremely constant like rain or wind.

Results

Results for the first HMM model and the test sounds are shown below in Figure 5. The error rate is calculated by:

$$ ER = \frac{\text{Model Incorrect Classifications per Class}}{\text{Model Correct Classifications per Class}} \times 100\% $$

<table>
<thead>
<tr>
<th></th>
<th>Detect Aircraft</th>
<th>Detect Not Aircraft</th>
<th>Error Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aircraft</td>
<td>6/9</td>
<td>3/9</td>
<td>33%</td>
</tr>
<tr>
<td>Cars</td>
<td>4/5</td>
<td>1/5</td>
<td>80%</td>
</tr>
<tr>
<td>Trains</td>
<td>2/5</td>
<td>3/5</td>
<td>40%</td>
</tr>
<tr>
<td>Trucks</td>
<td>2/3</td>
<td>1/3</td>
<td>67%</td>
</tr>
<tr>
<td>Dog Barks</td>
<td>0/3</td>
<td>3/3</td>
<td>0%</td>
</tr>
<tr>
<td>Rain</td>
<td>0/1</td>
<td>1/1</td>
<td>0%</td>
</tr>
<tr>
<td>Wind</td>
<td>0/1</td>
<td>1/1</td>
<td>0%</td>
</tr>
<tr>
<td>Thunder</td>
<td>1/1</td>
<td>0/1</td>
<td>100%</td>
</tr>
<tr>
<td>Music</td>
<td>0/1</td>
<td>1/1</td>
<td>0%</td>
</tr>
<tr>
<td>Totals</td>
<td></td>
<td></td>
<td>41%</td>
</tr>
</tbody>
</table>

Figure 5

Model Enhancements

In order to improve the model, first it was necessary to identify why certain sounds were being identified as aircraft noise. Upon reviewing the HMM, it was clear that any sound that had a wide band spectral structure, and slowly changed pitch was being classified as an aircraft. In addition, the small number of HMM states appeared to grossly represent the transition stage of the aircraft spectrum as the pitch changed. Three modifications were implemented in the second round of creating the aircraft noise detection system.
First Mel-Frequency Cepstral Coefficients were added as part of the feature vectors used to train and test the HMMs. The most important coefficient was \( \text{c0} \) or the log energy of the spectrum. When an aircraft fly over occurs, it is usually one of the loudest noises present in the environment at that time. The dull roar of the engine can be heard for miles, and should be represented in the model. The \( \text{c0} \) coefficient will provide the HMM a way to detect aircraft sounds based in part by the spectral energy that will slowly rise over time, then level off, and finally die away. A feature appending routine combined the feature vectors of both the LPC coefficients, and the MFCC coefficients. The MFCC feature extraction was achieved by using Malcolm Slaney's Toolkit. When the MFCC coefficients were added, I needed to reduce the windows size from 125 ms to 50 ms for both the LPC coefficients and MFCC coefficients in order to get the maximum benefit from the average log energy component.

Second, the HMM itself was lengthened to 7 states. This will provide for a better segmentation of the aircraft's spectrum features, and will provide a better representation of the structure when the aircraft noise changes pitch. The states that represent the approach and flyaway segment of the aircraft noise will not be affected by this change, as those states have a very high likelihood of transitioning back to itself rather than changing to the next state. However the intermediate states that transition more rapidly will benefit from additional states to provide better GMM classification.

Third, the segmentation of the training feature vectors was implemented by a routine that examined the \( \text{c0} \) MFCC coefficient, and determined logical separation points during the aircraft's fly over. As an aircraft approaches, the spectral energy remains relatively constant for a very long time, until the aircraft near the microphone vicinity. Then, the energy will quickly rise to it's peak level, remain constant, and then finally quickly fall to a level that remains relatively constant again. The routine would first identify the areas of the feature vectors that remain constant, and mark those two as part of the HMM's first and last state (not including the start and end state). Then the remaining feature vectors were split into 5 equal sections to approximately separate the transition stage into the 5 HMM states that make up the intermediate section.

The new aircraft detection system is shown in Figure 6 below.

![Figure 6](image-url)
Results

The results of the same set of test data on this new model is shown below in Figure 7. The error rate is calculated by:

\[ ER = \frac{\text{Model Incorrect Classifications per Class}}{\text{Model Correct Classifications per Class}} \times 100\% \]

<table>
<thead>
<tr>
<th></th>
<th>Detect Aircraft</th>
<th>Detect Not Aircraft</th>
<th>Error Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aircraft</td>
<td>7/9</td>
<td>2/9</td>
<td>22%</td>
</tr>
<tr>
<td>Cars</td>
<td>3/5</td>
<td>2/5</td>
<td>60%</td>
</tr>
<tr>
<td>Trains</td>
<td>1/5</td>
<td>4/5</td>
<td>20%</td>
</tr>
<tr>
<td>Trucks</td>
<td>1/3</td>
<td>2/3</td>
<td>33%</td>
</tr>
<tr>
<td>Dog Barks</td>
<td>0/3</td>
<td>3/3</td>
<td>0%</td>
</tr>
<tr>
<td>Rain</td>
<td>0/1</td>
<td>1/1</td>
<td>0%</td>
</tr>
<tr>
<td>Wind</td>
<td>0/1</td>
<td>1/1</td>
<td>0%</td>
</tr>
<tr>
<td>Thunder</td>
<td>0/1</td>
<td>1/1</td>
<td>0%</td>
</tr>
<tr>
<td>Music</td>
<td>0/1</td>
<td>1/1</td>
<td>0%</td>
</tr>
<tr>
<td>Totals</td>
<td></td>
<td></td>
<td>24%</td>
</tr>
</tbody>
</table>

Figure 7

Further Investigation

This model is designed to identify aircraft noises from various environmental sounds. The detection criteria is based solely on the log likelihood of the HMM model, and a decision is made that states if the sound came from an aircraft. A more sophisticated system would require a separate HMM for each class of environmental noise in order to more accurately predict the source of the sound. The system I designed consists of a single HMM that is trained on aircraft noise. If a separate HMM was trained for each additional source (train, car, dog bark, etc.), then the log likelihood could be compared between the various HMMs, and the decision would be made based on the HMM with the highest likelihood. This system was not designed as the complexity of creating many HMMs would fall outside the bounds of this project. Research projects have begun to create various HMMs for different environmental sounds [3], however this is a vast undertaking as the number of environmental sources that create sound is enormous. Complete ANR applications will require such techniques, however the benefits of a complete HMM environmental classifier will be tremendous.

A second area that was explored was the superpositioning of two environmental sources simultaneously. An example would be an aircraft fly over, and a loud dog barking. The system designed for this project is not capable of detecting the aircraft noise if a second noise source is simultaneously present. The HMM has a specific left-right structure, that will not allow for short term spectral changes that are inconsistent with aircraft noise. The log likelihood of such a noise source typically yielded results close to -Inf. While this project did not address changes needed to correctly handle these noise sources, it is clear that an HMM would need to be designed that is able to diverge onto a different HMM class model, and then simply merge back into the main Markov Chain. This would likely require the system as mentioned in the above paragraph that contains all the environmental noise HMMs, and then the global HMM would need the ability to
transition from any state in the Markov Chain, to the beginning of another HMM class Markov Chain. In addition, the final state of the secondary Markov Chain would require a transition state back to all the other states in the global HMM.

Conclusions

As can be seen from the results of this classification system, the model did reasonably well in detecting aircraft noise. It was difficult to distinguish aircraft noise from car noise, but the enhanced model improved the overall error rate. Upon completing this project, and reading the various research projects that have explored this same area, it became clear that LPC and MFCC coefficients were a limiting factor in classifying aircraft noise. Related research projects [5,7,8] into environmental noise detection have had success employing third-octave filtering in their system. The third-octave filter bank seems to due a better job at identifying pitch changes, and the long temporal characteristics of an environmental noise source.

A second area that various research projects are using is discrete HMMs with codebooks. Unlike the continuous HMM that I have used for this project, discrete HMMs are much easier to train, but require an additional step to identify each discrete feature vector with a specific codebook value. Due to the complexity of creating a codebook for the various aircraft feature vectors, a continuous HMM model was chosen for this project.

Reference