Classification of auditory signals from a combine harvester based on Mel-frequency Cepstral coefficients and machine learning

Gabriel Thomas[1], Avery Simundsson[2], Danny D. Mann[2] and Simone Balocco[3]

[1]*Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, MB, R3T 5V6 Canada.*
[2]*Department of Biosystems Engineering, University of Manitoba, Winnipeg, MB, R3T 5V6 Canada.*
[3]*Department of Mathematics and Computer Science, University of Barcelona, Gran Via 585, Barcelona, Spain.*

Corresponding Author: danny.mann@umanitoba.ca

ABSTRACT

As agricultural machinery moves into the digital era, significant developments in available technology will likely make autonomous farm vehicles more feasible, affordable, and desirable. One of the challenges of effective autonomous vehicle control specific to agriculture is the ability of the vehicle to interpret and adapt to constantly changing conditions. Auditory information is a primary indicator of changing conditions to an in-cab operator, particularly in situations such as detecting mechanical overload in a combine. This paper explores the potential for auditory information to be used in autonomous vehicle control. The sound was recorded at a sampling rate of 48 kHz near the straw chopper of a combine for three different operating modes during the same harvest day. Samples from each clip were segmented and analyzed to extract 31 audio features. Six different feature selection methods ranked the importance of each of the 31 features to identify the features that lead to accurate classification with a minimal number of calculations. These six rankings were assessed by Fagin's algorithm to yield two features (both mel-frequency cepstral coefficients). Twenty-five distinct machine learning classification methods were evaluated using these two features. Three of these classification methods reached 100% accuracy, and 9 classifiers exceeded an individual success rate of more than 99% using those same features. These feature extraction and classification steps took less than 1 s, assuring that such a classification system could be implemented in real-time.

KEYWORDS

Autonomous machines, remote supervision, auditory information, machine monitoring, mel-frequency cepstral features, machine learning, feature selection, Fagin's algorithm.

RÉSUMÉ

À mesure que les machines agricoles entrent dans l'ère numérique, les développements importants de la technologie offerte rendront probablement les véhicules agricoles autonomes plus pratiques, abordables et attrayants. L'un des défis du contrôle efficace des véhicules autonomes spécifiques à l'agriculture vient de la capacité du véhicule à interpréter et à s'adapter à des conditions qui changent constamment. Les informations auditives sont un indicateur primaire des conditions changeantes pour un opérateur en cabine, en particulier dans des situations comme la détection de la surcharge mécanique d'une moissonneuse-batteuse. Cet article explore le potentiel d'utilisation des informations auditives dans le contrôle des véhicules autonomes. Le son a été enregistré à une fréquence d'échantillonnage de 48 kHz près du broyeur de paille d'une moissonneuse-batteuse pour trois modes de fonctionnement différents au cours d'une même journée de récolte. Des échantillons de chaque enregistrement ont été segmentés et analysés pour extraire 31 caractéristiques audio. Six méthodes différentes de sélection des caractéristiques ont classé l'importance de chacune des 31 caractéristiques afin d'identifier celles qui permettent une classification précise avec un nombre minimal de calculs. Ces six classements ont été évalués par l'algorithme de Fagin pour obtenir deux caractéristiques (toutes deux des coefficients cepstraux à fréquences selon l'échelle de Mel). Vingt-cinq méthodes distinctes de classification par apprentissage machine ont été évaluées à l'aide de ces deux caractéristiques. Trois de ces méthodes de classification ont atteint une précision de 100 %, et neuf classificateurs ont dépassé un taux de réussite individuel de plus de 99 % en utilisant ces mêmes caractéristiques. Ces étapes d'extraction de caractéristiques et de classification ont pris moins d'une seconde, ce qui garantit qu'un tel système de classification peut être mis en œuvre en temps réel.

MOTS CLÉS

Machines autonomes, télésurveillance, information auditive, surveillance de machine, caractéristiques cepstraux à fréquences selon l'échelle de Mel, apprentissage machine, sélection de caractéristiques, algorithme de Fagin.

CITATION

Thomas, G., A. Simundsson, D.D. Mann and S. Balocco. 2021. **Classification of auditory signals from a combine harvester based on Mel-frequency Cepstral coefficients and machine learning.** Canadian Biosystems Engineering/Le génie des biosystèmes au Canada 63: 2.13-2.22. https://doi.org/10.7451/CBE.2021.63.2.13

INTRODUCTION

When seated inside the cab of the agricultural machine, the operator has access to various forms of sensory information (i.e., visual, auditory, tactile, olfactory) that can be used to supplement or complement information that machine designers have selected to display via the machine's instrument panel. Human operators must use a significant amount of intelligence and judgment to process these sensory cues rapidly and then react accordingly depending on the required vehicle operation or maneuver (Reid et al. 1999). In essence, the operator is an in-situ sensor capable of detecting sensory information used in decision-making.

It is recognized that machinery operators are often able to detect existing or impending problems from the changes in the sound produced by the mechanical components of the machine. Karimi et al. (2008) reported that the addition of auditory cues did not improve steering performance (in a simulated agricultural vehicle), perhaps because the steering is a purely visual task. However, auditory cues did improve the monitoring task. Donmez et al. (2009) investigated the use of sonification (continuous auditory alerts) during the control of unmanned aerial vehicles and found that visual information supported by sonifications yielded faster reaction times than visual information supported by discrete auditory signals. Sound can be a key part of the mechanical analysis in a conventional setting (Donmez et al. 2009). Many mechanics use sound as a tool for preliminary diagnosis, and even individuals with little mechanical aptitude understand that an unusual sound in a vehicle is a signal of an engine malady. Combine operators rely on sound as an indicator of over-capacity when threshing (Donmez et al. 2009). Even with extensive visual displays of information in modern combines, audible cues to overload conditions may allow for quicker response times (Donmez et al. 2009). Adjusting the concave clearance, cylinder speed, and the fan speed may be necessary to maximize harvest efficiency while minimizing losses and reducing seed damage. As conditions change throughout the day, these parameters should be monitored and adjusted. Automating these adjustments or even alerting an operator to the evolving conditions requiring adjustment would be a significant step forward in a fully autonomous harvesting machine. In the case of fully autonomous machines, it is unlikely that a human supervisor will be directly listening to each machine. However, it is important to determine whether real-time auditory information could contribute to the task of remotely supervising an autonomous agricultural machine.

This paper will explore the possibility of using the auditory information produced by a combine to detect changes in its mode of operation. If changes can be detected and accurately categorized using machine learning techniques, there may be a reason to incorporate such information into an automation interface for remote supervision of autonomous agricultural machines. The research described in this study expands upon the work previously presented at a conference by Simundsson et al. (2019). We have enriched the research by using more features and classification methods to develop a robust solution which can be implemented in a real-time system. The performance evaluation was verified by using a 10-fold cross-validation strategy.

LITERATURE REVIEW

Sound analysis & classification

Sound waves can be represented in several ways, but to analyze or manipulate them, they are represented in the form of an electrical quantity (Priemer 1990). Classification of music is an excellent example of processing audio signals for rapid identification. Music Information Retrieval (MIR) is a field of science that is becoming increasingly important as consumers become increasingly accustomed to tailored experiences in everything from movie selection to curated playlists of new songs. The goal of pattern recognition is to create a classifier that can analyze specific features of an item as its input and return a label or value indicating grouping to which the item belongs (Mahana and Singh 2015). The specific patterns that the algorithm can recognize are based on features of the signal. A feature is a distinctive measurement, transform, or structure component extracted from a pattern distinguished from a regular vector, to be used for classification. The purpose of feature extraction is to identify information that is most useful for determining the classification of the signal. The features are the inputs to the algorithm that are expected to predict the outcome. An example outside of music retrieval is classifying brain electrical activity from an electroencephalogram (EEG) signal to diagnose (Al-Fahoum and Al-Fraihat 2014).

Frequency analysis

Time-frequency analysis is commonly used to characterize phenomena such as vibration, music, and biomedical signals (Nisar et al. 2016). Fourier transforms are typically used to gain useful information from these phenomena, though this method ignores all time-related information. The ubiquitous use of the Fourier transform in signal processing, and analysis and unanimous acceptance as a valued function make it an obvious candidate for evaluation in this study.

Fourier transform, designed initially for continuous functions, can be numerically computed on digital signals using the Discrete Fourier Transform (DFT). The reduction of computational effort in FFT makes real-time DFT analysis practical in situations when it would otherwise be unfeasible. The FFT is a one-push algorithm that allows efficient implementation of the length N Periodogram (PG) of a signal x(n) calculated as:

$$X(k) = \left| \sum_{n=0}^{N-1} x(n) e^{-\frac{j2\pi kn}{N}} \right|^2 \quad \text{for k = 0,1, …, N-1} \tag{1}$$

where k is the frequency bins.

The FFT allows for simplification, reduced storage requirements, and reduced computational error. The FFT performs best when processing stationary signals compared to non-stationary signals (Al-Fahoum and Al-Fraihat 2014). It is particularly appropriate for narrowband signals (such as a sine wave) and has superior speed over almost all other methods for real-time applications. A stationary signal is one in which the statistical properties of a random process do not depend on the time index. The FFT is very suitable for real-time analysis because of its ubiquitous use in signal processing, error reduction, speed of processing, and low storage requirements.

Different from the Fourier transform, the Mel-Frequency Cepstrum (MFC) is based on a linear cosine transform of a log power spectrum on a nonlinear Mel frequency scale; Mel scale being a conversion from a linear frequency f to a logarithmic one to go from Hz to Mels as:

$$\text{Mel}\{f\}=2595\log(1+f/700) \quad (2)$$

Based on human perception and knowledge that the human ear can be seen as being composed of a bank of filters that are non-uniformly spaced with more filters concentrated in the lower bands than in the higher frequency ones, it has been noted that filters spaced linearly at low frequencies and logarithmically at the other end of the spectrum capture phonetical characteristics of human speech (Davis et al. 1980). As indicated in Equation 2, this logarithmic spacing makes Mel-frequency analysis a useful technique for speech processing.

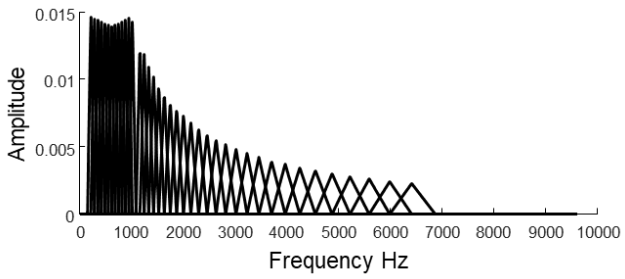
A useful frequency band starts at a low frequency, not necessarily zero, and is divided into M channels equidistant in the Mel frequency domain. Each channel is formed via a triangular frequency window. Consecutive channels are half overlapping. Figure 1 shows an example of such a filter bank.

The centre frequencies of the channels in terms of the FFT bin indices (k for the i^{th} channel) are calculated as:

$$f_{c_l} = Mel^{-1}\left\{Mel\{f_{start}\} + \frac{Mel\{\frac{f_s}{2}\}-Mel\{f_{start}\}}{M+1} l\right\},$$

$$l = 1,2,\dots,M \quad (3)$$

$$cbin_l = \text{round}\left\{\frac{f_{c_l}}{fs} N\right\} \quad (4)$$



The output of the Mel filter is the weighted sum of the FFT magnitude spectrum values in each band. Triangular, half-overlapped windowing is used as follows:

$$fbank_l = \sum_{k=cbin_{l-1}}^{cbin_l} \frac{k-cbin_{l-1}+1}{cbin_l-cbin_{l-1}+1} X(k)$$

$$+\sum_{k=cbin_{l-1}}^{cbin_{l+1}} 1-\frac{k-cbin_l+1}{cbin_{l+1}-cbin_l+1} X(k) \quad (5)$$

where l = 1, ..., M-1, cbin0 and cbinM denote the FFT bin indices corresponding to the starting frequency and half of the sampling frequency, respectively:

$$cbin_0 = \text{round}\left\{\frac{f_{start}}{fs} N\right\} \text{ and } cbin_M = N/2.$$

The output of the Mel filtering is subjected to a natural logarithm function:

$$f_l = \ln(fbank_l),\ l = 1,\dots,M-1 \quad (6)$$

Figure 2 shows an example of a periodogram and a mel-spectrogram for three different sounds of a harvester. An important application of the MFC is the extraction of coefficients known as Fourier Mel-Frequency Coefficients (MFCC), defined as:

$$C_i = \sum_{l=1}^{M} f_l cos\left(\frac{\pi i}{M}(l-0.5)\right),\ 0 \le i \le N_{MFCC} \quad (7)$$

where fl is defined as in (6), and NMFCC is the total number of coefficients.

The use of MFCCs to extract features from audio has been a technique used in audio classification (Li et al. 2013; Shen et al. 1999; Rong 2016). For example, using a treebagger as a classifier, feature vectors consisting of 10 MFCCs, the spectral centroid and spectral flux were successfully used for audio recognition by Li et al. (2013). Using a Mel-Spectrogram of 30 bands, MFCCs were used as audio features by Shen et al. (1999) to recognize Mandarin base syllables in quiet conditions and the presence of microphone variations. Another more recent example for audio classification of sounds using not just the MFCCs but also zero crossings and short time energies to

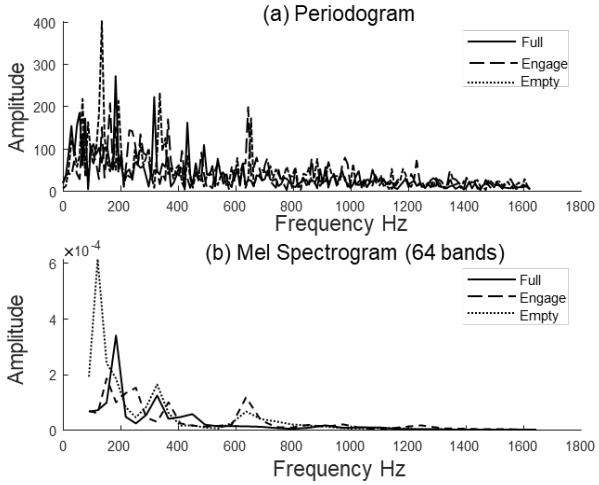


Fig. 7. Scatter plot of the features corresponding to the mel-frequency cepstrum coefficients C7 and C8.

form feature vectors was presented in Rong (2016) using a Support Vector Machine (SVM) as a classifier. Furthermore, an SVM and thirteen MFCCs extracted from audio segments were used to classify four types of medical pathologies: chronic laryngitis, cyst, Reinke edema and spasmodic dysphonia (Nawel et al. 2015). Detection of air drone sounds from other sounds in the sky (i.e., birds, airplanes, and thunderstorms) was also possible (Anwar et al. 2019). Thus, evidence suggests that this approach can be applied to areas other than speech recognition.

Research Objective

There has been a significant amount of research in recent years on using sound in classification and identification systems, particularly in music retrieval. There are various ways to do this, but one of the most straightforward and most robust is to take the Fast Fourier Transform of a signal and seek out specific features for identification. If characteristic features can be identified, they can create a classification system through various methods. Recent advances in computing power have made neural networks an excellent candidate for training classification models, mainly if more data is available.

This research aimed to determine whether the methodology used in other sound classification applications can also be used to classify machinery operations. Mainly, it will focus on the ability of machine learning techniques to correctly identify the operating condition of a combine with the goal of creating an analysis technique that could be used for real-time monitoring and control of an autonomous agricultural machine. In a study of driver perception response time, Olson and Sivak (1986) found that the average time for a person to sight an obstacle and apply the brake was 1.6 s for 95% of test drivers, regardless of age. A similar study showed that the 85th percentile of people have reaction times of 1.3-3.6 s depending on the driving conditions (night/day, moving vs stationary obstacle, etc.) (Triggs and Harris 1982). Though their studies took place on highways using personal vehicles, they cover a variety of driving conditions, and we can assume that reaction times to stimulus while operating farm machinery would be similar. Therefore, any system that can provide a real-time reaction time (i.e., time from sensing the issue to implementing a response mechanism) of less than 1 s can be considered faster than a human response and sufficient for a vehicle control system.

EXPERIMENTAL METHOD

Sound recordings were taken from harvest video collected during the 2017 canola harvest from a field near Selkirk, MB. The canola was harvested with an S680 John Deere combine, and video was captured with a GoPro Hero Session camera. The recordings were taken from the rear of the combine near the straw chopper (see Figure 3), always when the combine was in forward motion. All recordings were taken in the same field on the same day from the same machine. Sound, sampled at a rate of 48 kHz with AAC compression and automatic gain control, was lifted from the video and converted to .wav files for analysis.

Fig. 3. Location of the GoPro during operation from which sound was lifted.

Sound recordings were isolated into three different combine operating modes (classes):

- Mode 1: The combine's engine is running, but mechanized threshing is not engaged ("Empty")
- Mode 2: The combine's engine is running, and mechanized threshing is engaged with no actual threshing being performed ("Engaged")
- Mode 3: The combine's engine is running, and mechanized threshing is engaged and utilized at approximately 80% capacity ("Full")

A short clip of sound (30-36 s) was taken from each operating mode as a representative audio sampling for that operating mode. Each operating mode was assigned a class (1, 2, or 3). Each recording was considered a stationary signal, independent of time, and the combine ran in a steady state during each clip. Each clip was segmented into identically sized segments of 5000 samples (104.2 ms) to ensure that enough time was available to extract the features and perform the classification in less than the required 1 s metric.

AUDIO FEATURE EXTRACTION

The periodogram (PG) was calculated for each segment as in Equation 1 using 5000 audio samples, and the resulting PGs were grouped by class. Each block was also analyzed using the MCF using the same samples as explained in a previous section.

Thirty-one features were selected to build a classification model. Defining i as a frequency position and yi as the amplitude of the PG at that frequency, the features can be calculated as follows (the first eight features can be visualized in Figure 4):

- *f1*: The frequency bin of the first dominant peak (P1)
- *f2*: The frequency bin of the second dominant peak (P2)
- *f3*: The frequency bin of the third dominant peak (P3)

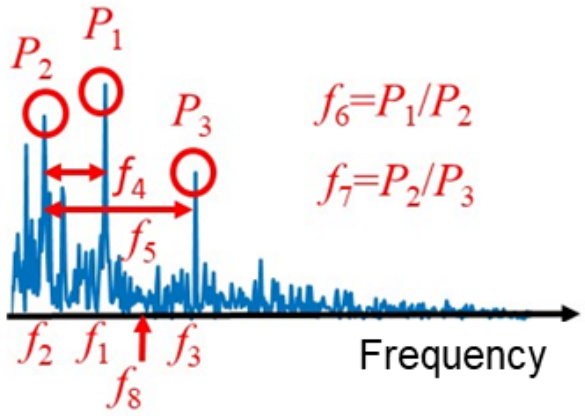


Fig. 4. Visual interpretation of the first six features.

- *f4*: The distance between the first and second dominant peaks
- *f5*: The distance between the second and third dominant peaks
- *f6*: The ratio of the magnitude of the first and second dominant peaks (P1/P2)
- f_7: The ratio of the magnitude of the first and second dominant peaks (P_2/P_3)
- f_8: The center of gravity (spectral centroid).

$$centroid = \frac{\sum_i i y_i}{\sum_i y_i} \quad (8)$$

- f_9: Coefficient of variation. Defined as the ratio of the standard deviation σ to the mean μ:

$$CV = \frac{\sigma}{\mu} \quad (9)$$

- f_{10}: Spectral spread

$$spread = \sqrt{\frac{\sum_i (i-centroid)^2}{\sum_i y_i}} \quad (10)$$

- f_{11}: Spectral skewness

$$skewness = \sqrt{\frac{\sum_i (i-centroid)^3 y_i}{(spread)^3 \sum_i y_i}} \quad (11)$$

- f12: Spectral kurtosis

$$kurtosis = \sqrt{\frac{\sum_i (i-centroid)^4 y_i}{(spread)^4 \sum_i y_i}} \quad (12)$$

- f_{13}: Spectral entropy

$$entropy = \frac{-\sum_{i=0}^{i=N/2} y_i \log(y_i)}{\log(N/2)}. \quad (13)$$

- f_{14}: Spectral flatness

$$flatness = \frac{\left(\prod_{i=0}^{i=N/2} y_i\right)^{N/2}}{\frac{1}{N/2}\sum_{i=0}^{N/2} y_i} \quad (14)$$

- f_{15}: Spectral crest

$$crest = \frac{\max(y_i \in [0, N/2])}{\frac{1}{N/2}\sum_{i=0}^{N/2} y_i} \quad (15)$$

- f_{16}: Spectral decrease

$$decrease = \frac{\sum_{i=1}^{N/2} \frac{y_i - y_0}{i-1}}{\sum_{i=1}^{N/2} y_i} \quad (16)$$

- f_{17} to f_{30}: Fourteen Mel-frequency cepstral coefficients calculated as in (7)
- f_{31}: Spectral slope

$$slope = \frac{\sum_{i=0}^{N/2}(f_i - \mu_f)(y_i - \mu_s)}{\sqrt{\sum_{i=0}^{N/2}(f_i - \mu_f)^2}} \quad (17)$$

where f_k is the frequency in Hz corresponding to bin k, μ_f is the mean frequency, and μs is the mean spectral value.

FEATURE SELECTION METHOD

Computing all the features in the previous section can be a time-consuming task. This is one good reason why only a few features should be used for a real-time classification system and highlights the importance of selecting the most valuable features. As Petri (2020) indicated, it is not just the amount of data but the data significance and usefulness for the application that must be considered. The importance of selecting the right type and number of input features in a classifier can be assessed mathematically. This is known as feature engineering, the systematic process that involves the selection of a subset of features that both speeds up the classification by performing fewer calculations and improves the performance of a machine learning algorithm (Duboue 2020). With this in mind, we used six different selection methods that ranked all 31 features.

Method 1: Univariate feature ranking for classification using chi-square tests (*FSCHI2*)

Chi-square tests evaluate the worth of a feature by computing the value of the chi-squared statistic with respect to a class. The initial assumption is that two features are unrelated, and it is measured by the chi-squared metric:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (18)$$

where, O_{ij} is the observed frequency and E_{ij} is the expected frequency. The greater this metric, the less likely is the assumption of the two features being unrelated (Novaković et al. 2011; Liu et al. 1995).

Method 2: Minimum redundancy maximum relevance (*FSMRMR*)

This method selects a subset of features having the most correlation with a class (relevance) and the least correlation between themselves (redundancy). The features are ranked according to the minimal-redundancy-maximal-relevance criteria. Relevance can be calculated by using the F-statistic (for continuous features) or mutual information (for discrete features), and redundancy can be calculated by using the Pearson correlation coefficient (for continuous features) or mutual information (for discrete features) (Ding et al. 2003; Radovic et al. 2017).

Method 3: Feature selection using neighbourhood component analysis for classification (*FSNCA*)

This method learns the feature importance by using a diagonal adaptation of neighbourhood component analysis (NCA) with regularization (Yang et al. 2012; Malan et al. 2019). Neighbourhood component analysis measures the Mahalanobis distance used in the KNN classification algorithm. A feature selection technique selects the best subset of features by maximizing an objective function that evaluates the average leave-one-out classification accuracy over the training data. The algorithm assesses a weighting vector w that corresponds to the feature vector x_i by optimizing the nearest neighbour learning classifier. A reference sample point x_j is selected for the sample xi from all the samples. The probability P_{ij} of x_j being chosen as a reference point for x_i from all the samples is higher depending on the closeness of the distance between the two samples.

Method 4 and 5: Rank importance of predictors using ReliefF algorithm using 5 and 10 nearest neighbours (*ReliefF5 ReliefF10*)

The algorithm randomly selects a sample x from the training set and searches for k nearest neighbour samples of the same class and k nearest neighbour samples of the non-similar classes. Using the Euclidean distance, the closest nearest neighbour samples from each class are selected. Each feature's relevant weight is assigned by comparing the interclass distance and interclass distance from the neighbour samples. This procedure is repeated on each feature sample, and each feature is assigned a weight. The algorithm penalizes the predictors that give different values to neighbours of the same class and rewards predictors that provide additional values to neighbours of different classes (Kononenko 1994; Robnik et al. 2003).

Method 6: Feature importance using a tree bagger for classification (TreeBagger)

When using a tree bagger for classification, compute the feature importance by permuting the values of features across every observation in the data set and measure how much worse the MSE becomes after the permutation (Breiman 2001). This process is repeated for each feature.

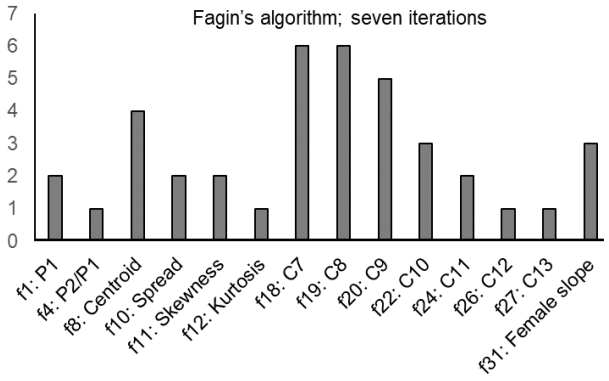


Fig. 5. Feature importance according to Fanig's algorithm. The curve represents the cumulative total.

COMBINING ALL THE RANKINGS

Method 1: Fagin's algorithm

Table 1 shows the rankings of the best 10 features (as described above), in descending order of importance, obtained using the six methods described in the previous section.

We used Fanig's algorithm (Wimmers et al. 1999) to combine all the different feature selection choices from Table 1 and to select a final subset of features to be used in a classifier. The method sequentially accesses all lists in parallel until there are k objects that have been seen in all lists, in our case, every feature selection method ranking. This simple and elegant technique keeps aggregating features in order of importance. As seen in Table 1, each row seen as an iteration of the algorithm shows that feature 19, the Mel-frequency cepstral coefficient C3, shows five times in the very first iteration (first row of the Table) and by the fourth iteration (fourth row of the table), it has been chosen by all the algorithms. By the seventh iteration, coefficient C2, feature 18, is present in all the rankings.

Figure 5 shows the final rankings after 7 iterations. Only two features (18 & 19) were present for all six calculation methods used. By contrast, four features (4, 12, 26 & 27) appeared only once.

Table 1. Rankings of the best 10 features in descending order of importance.

Iteration	Method 1: FSCHI2	Method 2: FSMRMR	Method 3: FSNCA	Method 4: ReliefF5	Method 5: ReliefF10	Method 6: TreeBagger
1	**19**	**19**	8	**19**	**19**	**19**
2	**18**	8	10	20	20	20
3	20	11	12	22	22	31
4	8	27	**19**	**18**	**18**	8
5	1	4	**18**	24	24	**18**
6	31	31	11	14	26	22
7	10	**18**	20	12	17	1
8	29	20	27	26	12	28
9	22	29	23	10	11	10
10	23	22	24	17	8	6

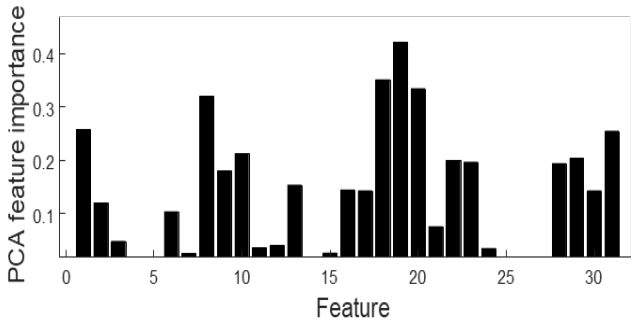


Fig. 6. Feature importance using PCA analysis.

Method 2: First Principal Component Projection Score (FPSPS)

Since some features are derived from functions of others, a final ranking considering Principal Component Analysis (PCA) was also considered. Let's define each entry on $X = \{x_1, x_2, \ldots, x_m\}$ as one of the 5000 audio samples blocks extracted from the three audio files. Let $F = \{f_1, f_2, \ldots, f_{31}\}$ be the set of 31 features extracted from each block so that $x_{ij} = f_j(x_i)$ denotes the j^{th} feature of the i^{th} block. Let $FS = \{\text{FSCHI2}, \text{FSMRMR}, \text{FSNCA}, \text{ReliefF5}, \text{ReliefF10}, \text{TreeBagg}$ be the set of the different feature selection methods that performed the ranking on Table 1.

To combine these different scorings, the FPSPS method considers each feature f_j as an entry on X and the scoring lists of the different feature selection methods becomes the features (Filchenkov et al. 2015). To reduce these scoring options to a final one, this can be seen as a dimension reduction problem in which the first principal component using PCA yields a final scoring. Results using this method yields the following importance for the 31 features, as shown in Figure 6. As can be seen, features 19 and 18 are the most important ones agreeing with the results obtained using Fagin's.

RESULTS

From Figures 5 and 6, we can identify two features from the Mel-frequency cepstrum ones, coefficients C7 and C8, as the most important features for classification. Figure 7 is a scatter plot of those two features where we can see how they cluster nicely for the three sound cases.

We used 25 different classifiers using 10-fold cross-validation and selected the two and three mel-frequency cepstrum features chosen in the previous sections as inputs. Table 2 summarizes the results.

Using the two features identified by the final rankings yielded 100% accuracy for three of the 25 classification methods. We used an order 3 polynomial kernel for the SVM Cubic classifier. For the SVM Medium Gaussian, the kernel used was Gaussian with a kernel scale of 1.4 (Christianini et al. 2000). For the Neural Network, one hidden layer with ten nodes and scaled conjugate gradient backpropagation was used for training.

Table 3 lists the execution times for feature extraction and classification. Processing was completed using an Asus

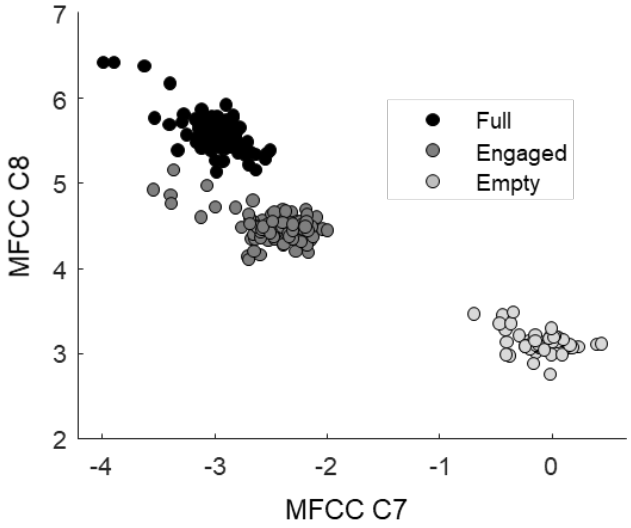


Fig. 7. Scatter plot of the features corresponding to the mel-frequency cepstrum coefficients C7 and C8.

laptop with a 64 bit Intel i7 CPU @ 2.6 GHz and 16 GB of ram, and the Matlab platform version 2020a.

Looking at Figure 2, the frequency content of these audio signals does not exceed 1 kHz, opening the possibility of sampling at a lower rate, in this case, 2 kHz. This reduction of sampling frequency increments the time between samples. It allows for implementing the solution in a computer platform that is not as fast as the one used, as

Table 2. Results of 10-fold cross-validation for two and three mel-frequency cepstrum features.

	Features used	
Classification	C_7,C_8	C_7,C_8,C_9
Tree (fine)	98.6	98.8
Tree (medium)	98.8	98.8
Tree (coarse)	98.8	98.8
Linear discriminant	99.6	99.2
Quadratic discriminant	99.6	99.2
Naïve Bayes (Gaussian)	98.3	99.6
Naïve Bayes (Kernel)	98.8	98.3
SVM (linear)	99.6	99.2
SVM (quadratic)	99.6	99.2
SVM (cubic)	**100**	99.2
SVM (fine Gaussian)	92.9	98.3
SVM (medium Gaussian)	**100**	99.2
SVM (coarse Gaussian)	98.8	98.8
KNN (fine)	99.2	98.8
KNN (medium)	99.6	98.8
KNN (coarse)	77.2	77.2
KNN (cosine)	98.3	98.3
KNN (cubic)	99.6	98.8
KNN (weighted)	99.6	98.8
Ensemble (boosted trees)	40.2	40.2
Ensemble (bagged trees)	99.6	99.2
Ensemble (Subspace discriminant)	98.8	98.8
Ensemble (Subspace KNN)	92.9	99.2
Ensemble (RUSBoosted trees)	81.7	82.2
Neural network	**100**	99.6

Table 3. Execution times for feature extraction and classification.

Step	Execution time (s)
Feature extraction	0.00054
SVM Cubic classification	0.0063
SVM Medium Gaussian classification	0.0047
Neural network classification	0.0077

the same number of operations will be needed to be completed in a longer execution time. It is expected that the ranking of the Mel coefficients will differ, as the filter bank shown in Figure 1 will cover the frequencies between 0 to 10 kHz and the distribution of the sinusoidal components of the audio signal would spread differently in the filter bank. Having this in mind, the audio signals were down sampled to have a 2 kHz sampling frequency effectively, computed the Mel coefficients and ranked the importance of the 14 coefficients using the FPSPS method. Figure 8 shows the results. As expected, the same coefficients are chosen when the ranking of the Mel coefficients is done at the 48 kHz sampling frequency but differs for the 2 kHz case.

Figure 9 shows the scatter plot of coefficients C_1 and C_{10} when sampling at 2 kHz. As can be seen, 100% accuracies are achieved using the same classifiers highlighted in Table 2.

CONCLUSIONS

In this paper, a method for creating a real-time classifier for classifying sounds from an operating agricultural machine (combine harvester) has been presented. The technique used the mel-frequency cepstral coefficients for feature extraction. Using six different feature selection methods and combining the feature rankings using Fanig's algorithm, two features were identified as the most important features for classification. Feature vectors were calculated for three different operating modes of the combine harvester: 1) Engine running with no threshing, 2) Engine running and threshing engaged but not loaded, and 3) Engine running and threshing engaged at approximately 80% capacity. Using the two features identified by Fanig's algorithm yielded 100% accuracy for three of the 25 classification methods.

At audio sampling rates of 48 kHz and 2 kHz and a segment size of 5000 samples, the data collection and execution times of the feature extraction and classification steps are sufficiently fast that future implementation of these results could be an automated classification system that produces a final decision that is based on five consecutive classifications (and taking the majority decision based on those five consecutive classifications). Thus, this method can successfully be used for real-time analysis and control of a combine harvester. The simplicity of using only one type of feature (Mel coefficients) and that of the chosen classifier (SVM) allow us to envision a low-cost hardware implementation. With further refinement, this information may be used to estimate and adjust the loading of the threshing system for optimal performance, reducing downtime and mechanical damage while increasing harvest efficiency.

RECOMMENDATIONS AND FUTURE WORK

Further work in this area should include a more significant number of operating modes. It would be useful to identify events that may cause machinery damage or harvest delays, such as overloading the feeder house, which can cause lengthy delays due to the required shutdown and manual extraction of material. Increasing the dimensionality of the classifier would allow it to provide more information to a

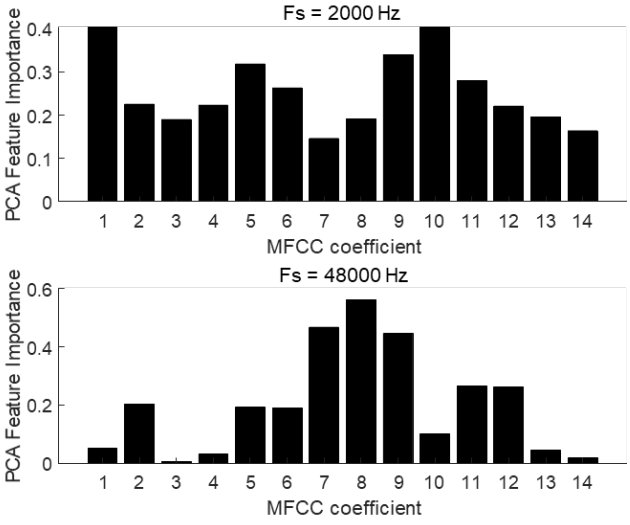


Fig. 8. Top: Importance of the Mel coefficients when sampling at 2 kHz using FPSPS. Bottom: The same coefficients are chosen by the FPSPS method when keeping the 48 kHz rate.

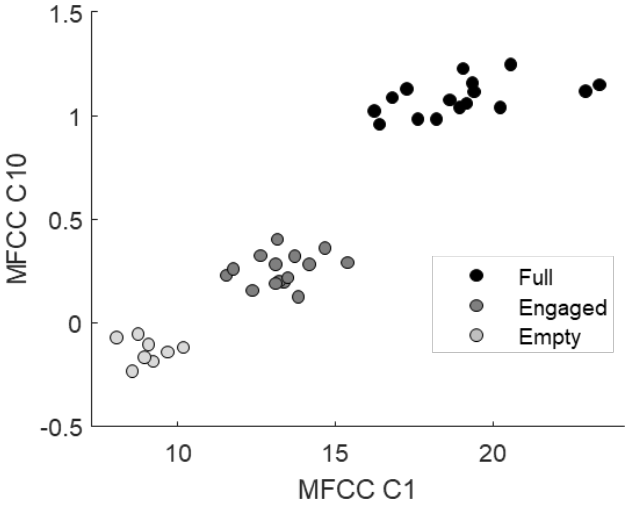


Fig. 9. Top: Scatter plot of coefficients C_1, and C_{10} for 2 kHz sampling.

remote operator, an automatic controller, or provide valuable data for diagnostics in machinery maintenance and repair. Future work can also include using spectrograms to train a classifier, different features, or combinations of features to optimally identify different operating modes or sensor selection/placement to optimize data collection and reduce this system's initial and maintenance costs.

Automatic feature extraction and selection could also be used with no need to visually inspect the absolute FFT for each class and manually select features. Matlab, Weka, and other machine learning tools (scikit-learn.org) offer this functionality and would likely provide an optimal set of features for efficient classification.

ACKNOWLEDGEMENT

The authors would like to acknowledge the support of technical staff at the Prairie Agricultural Machinery Institute (PAMI) in Portage la Prairie, MB, who assisted with data collection.

REFERENCES

Al-Fahoum, A.S., and A. A. Al-Fraihat. 2014. Methods of EEG signal features extraction using linear analysis in frequency and time-frequency domains. *ISRN Neuroscience* Volume 2014, Article ID 730218, 7 pp. https://doi.org/10.1155/2014/730218

Anwar, M. Z., Z. Kaleem and A. Jamalipour. 2019. Machine learning inspired sound-based amateur drone detection for public safety applications. *IEEE Transactions on Vehicular Technology* 68(3):2526-2534. https://doi.org/10.1109/TVT.2019.2893615

Bechar, A. and C. Vigneault. 2016. Agricultural robots for field operations: concepts and components. *Biosystems Engineering* 149:94-111.
https://doi.org/10.1016/j.biosystemseng.2016.06.014

Blackmore, B.S., H. Have and S. Fountas. 2002. A proposed system architecture to enable behavioural control of an autonomous tractor. Automation Technology for Off-Road Equipment. ed. Q. Zhang. 2950 Niles Road, St. Joseph, MI 49085-9659, USA. pp.13- 23.

Breiman, L. 2001. Random forests. *Machine Learning* 45:5-32. https://doi.org/10.1023/A:1010933404324

Christianini, N. and J. Shawe-Taylor. 2000. An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge, UK: Cambridge University Press.
https://doi.org/10.1017/CBO9780511801389

Davis, S.B. and P. Mermelstein. 1980. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 28(4):357-366.
https://doi.org/10.1109/TASSP.1980.1163420

Ding, C. and H. Peng. 2003. Minimum redundancy feature selection from microarray gene expression data. IEEE Bioinformatics Conference, Stanford, CA, pp. 523-528.

Donmez, B., M.L. Cummings and H.D. Graham. 2009. Auditory decision aiding in supervisory control of multiple unmanned aerial vehicles. *Human Factors* 51(5):718-729.
https://doi.org/10.1177/0018720809347106

Duboue P. 2020. The Art of Feature Engineering, Cambridge University Press.
https://doi.org/10.1017/9781108671682

Edet, U. and D.D. Mann. 2020. Visual information requirements for remotely supervised autonomous agricultural machines. *Applied Sciences* 10,2794. https://doi.org/10.3390/app10082794

Filchenkov, A., Vladislav Dolganov, V. and I. Smetannikov. 2015. PCA-based algorithm for constructing ensembles of feature ranking filters. Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, Bruges Belgium, pp. 201-206.

Johnson, D.A, D.J. Naffin, J.S. Puhalla, J. Sanchez and C.K. Wellington. 2009. Development and implementation of a team of robotic tractors for autonomous peat moss harvesting. *Journal of Field Robotics* 26(6-7):549-571. https://doi.org/10.1002/rob.20297

Karimi, D., T.A. Mondor and D.D. Mann. 2008. Application of auditory signals to the operation of an agricultural vehicle: results of pilot testing. *Journal of Agricultural Safety and Health* 14(1):71-78.
https://doi.org/10.13031/2013.24124

Kononenko, I. 1994. Estimating attributes: analysis and extensions of RELIEF. European Conference on Machine Learning, pp. 171-182.
https://doi.org/10.1007/3-540-57868-4_57

Li, D., J. Tam, and D. Toub. 2013. Auditory scene classification using machine learning techniques. *IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events.* [Online] Available: http://c4dm.eecs.qmul.ac.uk/sceneseventschallenge/abstracts/SC/LTT.pdf.

Liu, H. and R. Setiono. 1995. Chi2: feature selection and discretization of numeric attributes. *IEEE International Conference on Tools with Artificial Intelligence,* pp. 388-39.

Mahana, P. and G. Singh. 2015. Comparative analysis of machine learning algorithms for audio signals classification. International *Journal of Computer Science and Network Security* 15(6):49-55.

Malan, N. S. and S. Sharma. 2019. Feature selection using regularized neighbourhood component analysis to enhance the classification performance of motor imagery signals. *Computers in Biology and Medicine* 107:118-126.
https://doi.org/10.1016/j.compbiomed.2019.02.009

Moorehead, S., C. Ackerman, D. Smith, J. Hoffman and C. Wellington. 2009. Supervisory control of multiple tractors in an orchard environment, in 4th IFAC

International Workshop on Bio-Robotics, Information Technology and Intelligent Control for Bioproduction Systems, 2009.

Nawel, S. and A. Cherif. 2015. Dimensionality reduction for voice disorders identification system based on mel frequency cepstral coefficients and support vector machine. *7th International Conference on Modelling, Identification and Control*, Sousse, pp. 1-6.

Nisar, S., O.U. Khan and M. Tariq. 2016. An efficient adaptive window size selection method for improving spectrogram visualization. *Computational Intelligence and Neuroscience* Volume 2016, Article ID 6172453, 13 pages. https://doi.org/10.1155/2016/6172453

Novaković J., P. Strbac and D. Bulatović. 2011. Toward optimal feature selection using ranking methods and classification algorithms. Yugoslav *Journal of Operations Research* 21:119-135.
https://doi.org/10.2298/YJOR1101119N

Olson, P.L. and M. Sivak. 1986. Perception-response time to unexpected roadway hazards. *Human Factors* 28(1):91-96.
https://doi.org/10.1177/001872088602800110

Panfilov, I. and D.D. Mann. 2018. The importance of real-time visual information for the remote supervision of an autonomous agricultural machine. *Canadian Biosystems Engineering* 60:2.11-2.18.
https://doi.org/10.7451/CBE.2018.60.2.11

Petri. D. 2020. Big data, dataism and measurement. *IEEE Instrumentation & Measurement Magazine* 23(3):32-34. https://doi.org/10.1109/MIM.2020.9082796

Priemer, R. 1990. Signals and Signal Processing. In Introductory Signal Processing, by Roland Priemer, 1-10. Chicago, IL: World Scientific Publishing Company. https://doi.org/10.1142/0864

Radovic, M., M. Ghalwash, N. Filipovic et al. 2017. Minimum redundancy maximum relevance feature selection approach for temporal gene expression data. *BMC Bioinformatics* 18.
https://doi.org/10.1186/s12859-016-1423-9

Reid, J.F., Q. Zhang and N. Noguchi. 1999. Agricultural vehicle navigation using multiple guidance sensors. UILU-ENG-99-7013.

Robnik-Šikonja, M. and I. Kononenko. 2003. Neighborhood component feature selection: theoretical and empirical analysis of ReliefF and RReliefF. *Machine Learning* 53:23-69.
https://doi.org/10.1023/A:1025667309714

Rong. F. 2016. Audio classification method based on machine learning. *International Conference on Intelligent Transportation, Big Data & Smart City* (ICITBS), Changsha, pp. 81-84.
https://doi.org/10.1109/ICITBS.2016.98

scikit-learn.org. https://scikit-learn.org/stable/index.html

Shen, J. and W. L. Hwang. 1999. New temporal features for robust speech recognition with emphasis on microphone variations. *Computer Speech & Language* 13(1):65-78. https://doi.org/10.1006/csla.1998.0050

Simundsson, A., D.D. Mann and G. Thomas. 2019. A neural network to classify auditory signals for use in autonomous harvester control systems. CSBE/SCGAB 2019 Annual Conference, 14-17 July 2019, Vancouver. Paper No. CSBE19-163.

Stentz, A., C. Dima, C. Wellington, H. Herman & D. Stager. 2002. A system for semi-autonomous tractor operations. *Autonomous Robots* 13:87-104.
https://doi.org/10.1023/A:1015634322857

Triggs, T.J. and W.G. Harris. 1982. Reaction Time of Drivers to Road Stimuli. Victoria: Monash University.

Wimmers, E.L., L.M. Haas, M.T. Roth and C. Braendli. 1999. Using Fagin's algorithm for merging ranked results in multimedia middleware. In *Proceedings Fourth IFCIS International Conference on Cooperative Information Systems,* Scotland UK, pp. 267-278.
https://doi.org/10.1109/COOPIS.1999.792176

Yang, W., K. Wang and W. Zuo. 2012. Neighborhood component feature selection for high-dimensional data. *Journal of Computers* 7:161-168.
https://doi.org/10.4304/jcp.7.1.161-168