Computers and Electronics in Agriculture 159 (2019) 50–58

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Tree size estimation from a feller-buncher's cutting sound

Pengmin Pan[a,*], Timothy McDonald[b]

[a] Department of Biosystems Engineering, Auburn University, 308 Corley Building, Auburn, AL 36849, USA
[b] Department of Biosystems Engineering, Auburn University, 224 Corley Building, Auburn, AL 36849, USA

1. Introduction

It is common in cut-to-length (CTL) harvesting systems to collect a great deal of stem-level size and form data (Skogforsk, 2015) during harvest and these data can be used to analyze, for example, harvesting productivity (Olivera et al., 2016) or to make bucking decisions. A similar approach has not been developed, however, in tree-length timber harvesting systems common in the US South, but would be useful in understanding site-specific factors affecting machine performance. The limiting technology in development of stem characterization systems for tree-length harvesting is a simple and reliable means of measuring size while felling.

From a data collection perspective, CTL harvesting equipment has the advantage of handling, at the stump, each stem along most of its length and can therefore derive size, form, and location data while performing its normal duties (Leitner et al., 2014). Feller-bunchers, however, handle multiple stems only at the base, leaving little opportunity to make measurements on size at the stump. There is an opportunity, however, to take an indirect measure of stem diameter at breast height (DBH) while felling.

Several technologies have been proposed for measuring stem size in tree-length harvesting. Among others, these methods have included computer vision (Kan et al., 2008; Thamrin et al., 2013), ultrasonic sensing (Escolà et al., 2011), and optical sensing (McDonald et al., 2003). These technologies, however, required specific stand conditions to achieve good accuracy, mainly the absence of an understory that might interfere with measurements, or were complex and required expensive hardware to implement. Another means of extracting information on stem size that would not require a great deal of additional expensive technology would be the use of sound. Most tree-length harvesting in the US South is done using a feller-buncher equipped with a circular saw and this combination produces a very distinct sound while cutting that is easily identified by a human listener. Sound acquisition is inexpensive, robust, and its application might be less susceptible to interference from competition. It should, therefore, be suitable as a diameter sensing technology given acceptable levels of accuracy could be achieved.

Sound classification has been applied successfully in multiple fields, for example, speech (Cho and Park, 2016; Peng et al., 2016; Yogesh et al., 2017) or manners recognition (Ferreira and Alarcão, 2016; Grozdić et al., 2017), disease diagnoses (Zheng and Guo, 2017; Azmy, 2015; Mayorga et al., 2016; Vandermeulen et al., 2016), animal sound classification (Li and Wu, 2015; Lin et al., 2015; González-Hernández et al., 2017), object recognition (Kojima et al., 2016; Korucu et al., 2016), vehicle engine fault detection (Anami and Pagi, 2013; Pagi et al., 2015; Huang et al., 2017), and environmental event recognition (Grézl and Černocky, 2009; Chang and Chang, 2013). However, reports on using sound from a working environment to quantify an object's properties (such as grain moisture measurement (Amoodeh et al., 2006), size and shape by human listener (Grassi, 2005)) are limited, especially in automatically detecting object size.

Feller-bunchers essentially drive through a tree to cut it, and, if machine's travel speeds through trees were uniform, the cut duration and stem size will be highly correlated in tree-length timber harvesting. The hypothesis of this study is, that the sound of cutting is heard by human as substantially different form sound of other machine activities, mainly moving, therefore it may be also recognized by the electronic device/application. Therefore the objective of this research was to: (1) measure cutting time based on characteristic sound heard by human, (2) measure cutting time by electronic device, and finally, (3) correlate cutting time with tree diameter, where characteristic sound of cutting is recorded and sound time measured electronically.

2. Material and methods

2.1. Data collection

279 loblolly pines were randomly selected from two logging sites (181 from site one and 98 from site two) in central Alabama, U.S. Site one was more brushy and sloping, compared with site two, which would potentially add more noise to the audio data (mostly to the background sounds). Each tree was numbered using a color code applied using ribbon tape (yellow, red, blue, and orange, 1–4 rings), examples of which could be found in Fig. 1a (orange 2 & 3). Their DBHs were manually measured with a calibrated diameter tape (d-tape, with an accuracy of ± 0.4 cm (Andersson and Dyson, 2001)). The mean and standard deviation of measured DBHs are 27.50 cm and 7.40 cm, respectively, and the distribution of DBH values is plotted in Fig. 1b. All

* Corresponding author.
E-mail addresses: pzp0014@auburn.edu (P. Pan), mcdontp@auburn.edu (T. McDonald).

https://doi.org/10.1016/j.compag.2019.02.021
Received 12 March 2018; Received in revised form 15 February 2019; Accepted 19 February 2019
Available online 01 March 2019
0168-1699/

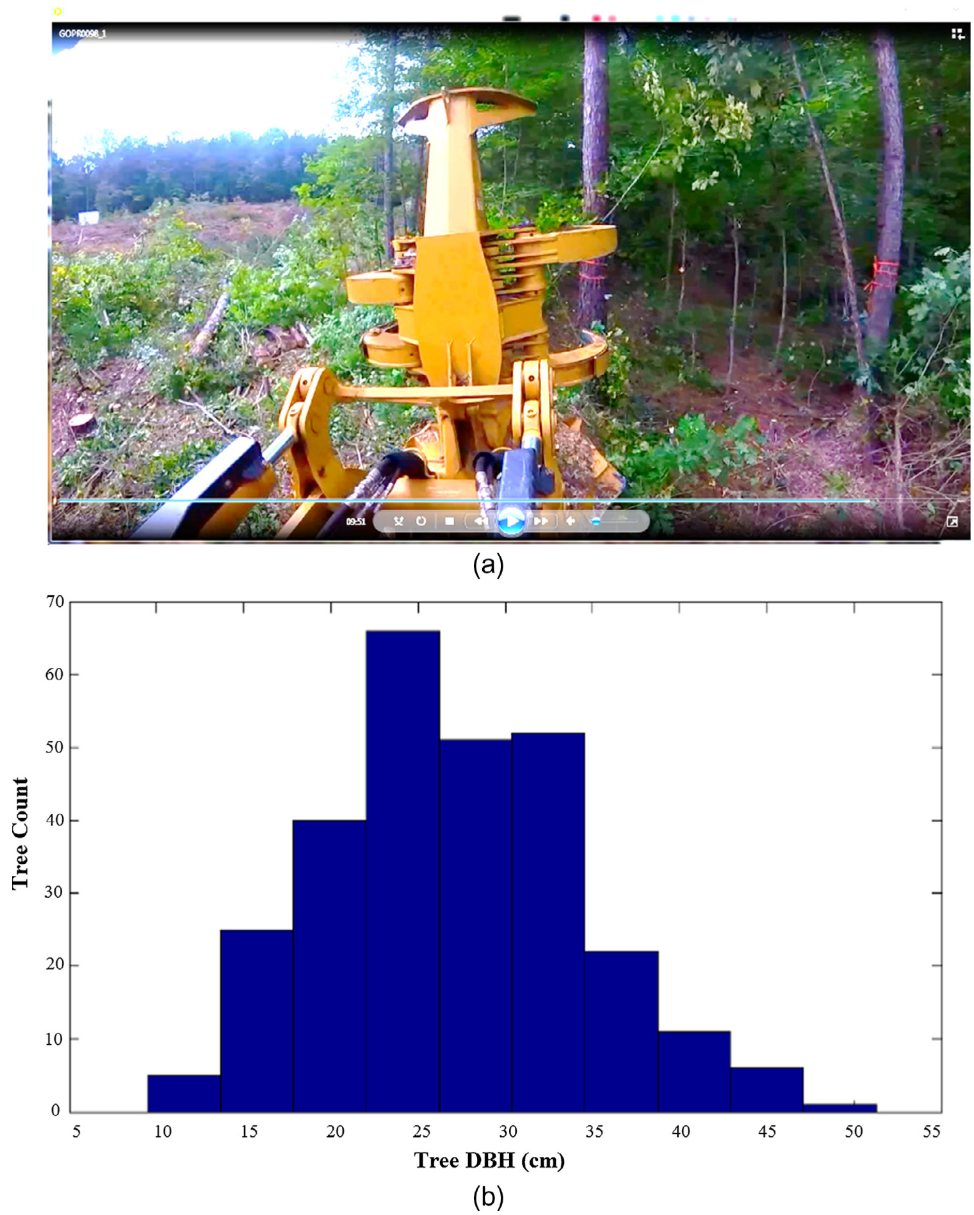

Fig. 1. (a) An example of flagged trees from recorded video, (b) Histogram of measured tree.

trees were harvested using two different models and ages of TigerCat (Tigercat International Inc., East Brantford, Ontario, Canada) feller-bunchers, one machine per site. Videos of the timber harvest were made using a GoPro (GoPro, Inc., San Mateo, California, U.S.) camera mounted on the right side of the feller-buncher's cab. The video recordings afforded a forward view sufficient to identify most marked trees in order to associate sounds with diameters. Audio (16-bit, 48 kHz) was extracted from the video recordings using Matlab software (*audioread* function, The MathWorks, Inc., Natick, Massachusetts, U.S.).

2.2. Manual cut duration (MD) calculation

The collected video/audio data was a mixture of cut sounds from marked and unmeasured trees. To simplify the targeted duration calculation, the marked tree audios were manually clipped and saved as fragments of a constant length. The initial review of the audio records indicated the longest cutting duration in the dataset was less than 2.1 s (for a 20 in. DBH tree). Each portion of the audio identifiably associated with a tree cut and a DBH was, therefore, extracted from the continuous audio stream into a 3-s clip. No effort was made to center the cut event within the 3-s window, but for each one there was some period of not-cutting background sound before, and after, the saw came in contact with the tree. The sound clips examples can be found in the linked Mendeley Data and the full data with DBHs are available at DOI: https://doi.org//10.17632/d8zvkyf4h2.1.

Each audio clip was reviewed by a single individual to establish their best estimate of the beginning and end of audible saw contact with the tree, the difference of which was the definition of cut duration used in this study. This duration was used as a reference and compared to automated methods of extracting saw contact duration. The clip was played back using Media Player Classic (available at: https://mpc-hc.org/) simultaneously showing the time record of the track with resolution of one millisecond, and variable playback speed. In most cases, the time record and sound during saw contact were clearly distinct, primarily because of a large difference in sound intensity.

2.3. Automated cut duration (AD) measurement

Visually distinct regions were evident in spectrogram images of the feller sound data associated with cut events (Fig. 2). Those regions

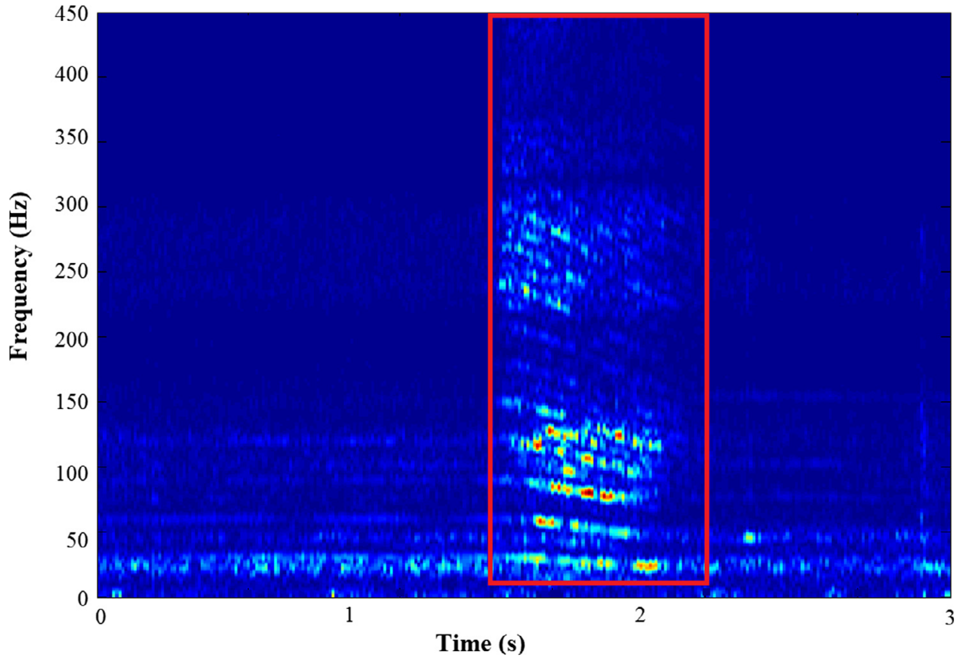

Fig. 2. Spectrogram example of sound clip including a 21.8 cm DBH tree cut as in red box. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

showed a higher degree of patterning in a specific range of the audio spectrum, between 50 and 450 Hz. A pattern recognition-based image recognition method was applied to classify sound clips into regions of 'cut' and 'background', the Local Binary Pattern (LBP). This classifier is an image classification tool whose features are scale and rotation invariant measures of textural characteristics in a gray-scale image (Ojala et al., 2002). They have been used successfully in a wide array of object recognition and pattern classification applications, including facial recognition (Akariman et al., 2015; Faudzi and Yahya, 2014) and detection of other objects (Heikkilä et al., 2009; Satpathy et al., 2014; Heng et al., 2012).

Neural Networks (NN) are a powerful tool that has been widely used in sound classification, for example in distinguishing the source of sounds at a metro train station (Grézl and Černocky, 2009), from real life audio (Valenti et al., 2017), for identifying hitting or scratching among objects (Owens et al., 2016), *etc*. The method requires a large amount of training data to adjust weights of neurons and edges, and is very suitable for use in conjunction with LBP features to classify the sound spectrograms. In this study, the LBP features derived from the reduced (50–450 Hz) spectrogram were identified as a reasonably sensitive feature for use as input to a NN classifier for discriminating cut/not-cut sounds.

The steps outlined below, and illustrated in Fig. 4, were used to create the LBP-based NN classifier for felling sounds.

(1) Raw sound data (Fig. 3a) were separated into 3-s clips, each containing a cut sound, a total of 279 (Fig. 3b). The clips were identical to those used in the MD experiments.
(2) For each clip, a log spectrogram was calculated using a fast Fourier transform (FFT) with window length of 1024 points, zero-padded to a total length of 4096, and overlap of 512 points. Data below 50 Hz and above 450 Hz were discarded. The result for each clip was a 300 × 280 matrix (image) of gray values (Fig. 3c, in pseudo-color).
(3) A random subset of the 279 available clips (total 200) was partitioned manually into cutting and not-cutting zones to generate training data for the NN classifier. Depending on the cut duration, there were sometimes multiple cutting and not-cutting zones extracted from each clip. Zones were of variable duration.
(4) For each zone, an LBP feature vector was calculated using the Matlab '*extractLBPFeatures*' function. The process resulted in 266 and 532 cutting/not-cutting feature vectors, respectively (Fig. 3d). This database of feature vectors was split randomly into training, validation, and testing groups with 85/5/10 percent of the total, respectively.
(5) A neural net implementing conjugate gradient back-propagation with five hidden neurons (2 layers) was constructed and trained using the 226/452 cutting/not-cutting LBP feature vectors of the training sub-group (Matlab, pattern recognition tool). After training, the NN model achieved a 99% correct classification rate (7 incorrectly classified) on the calibration dataset, and 100%/99% (1 misclassified) on the validation and testing datasets, respectively.

The trained neural net identified regions (0.1-s in length) of cutting and non-cutting and was applied on the 200 3-s sound clips from the training set. The longest contiguous period in any single clip identified as being of cutting class was used as the cut duration. These data were then regressed with stem DBH to generate a stem size prediction model. The procedure was then repeated on the validation set of clips (79 total) to verify the prediction model's performance.

2.4. Method evaluation

A linear regression method at the confidence level of 95% was used to test hypotheses: MD is highly correlated with tree DBH, as well as with AD. The parameter used was the slope and correlation coefficient (R^2). After the hypotheses validation, the LBP_NN model's performance was evaluated by linear regression (AD VS. DBH in calibration set) and root mean square errors (RMSE) of both model and prediction sets. The calculation of RMSE was:

$$RMSE = \sqrt{\frac{\sum_1^n (DBH_{estimated} - DBH_{measured})^2}{n}} \tag{1}$$

where n was 200 and 79 for model and prediction sets, respectively.

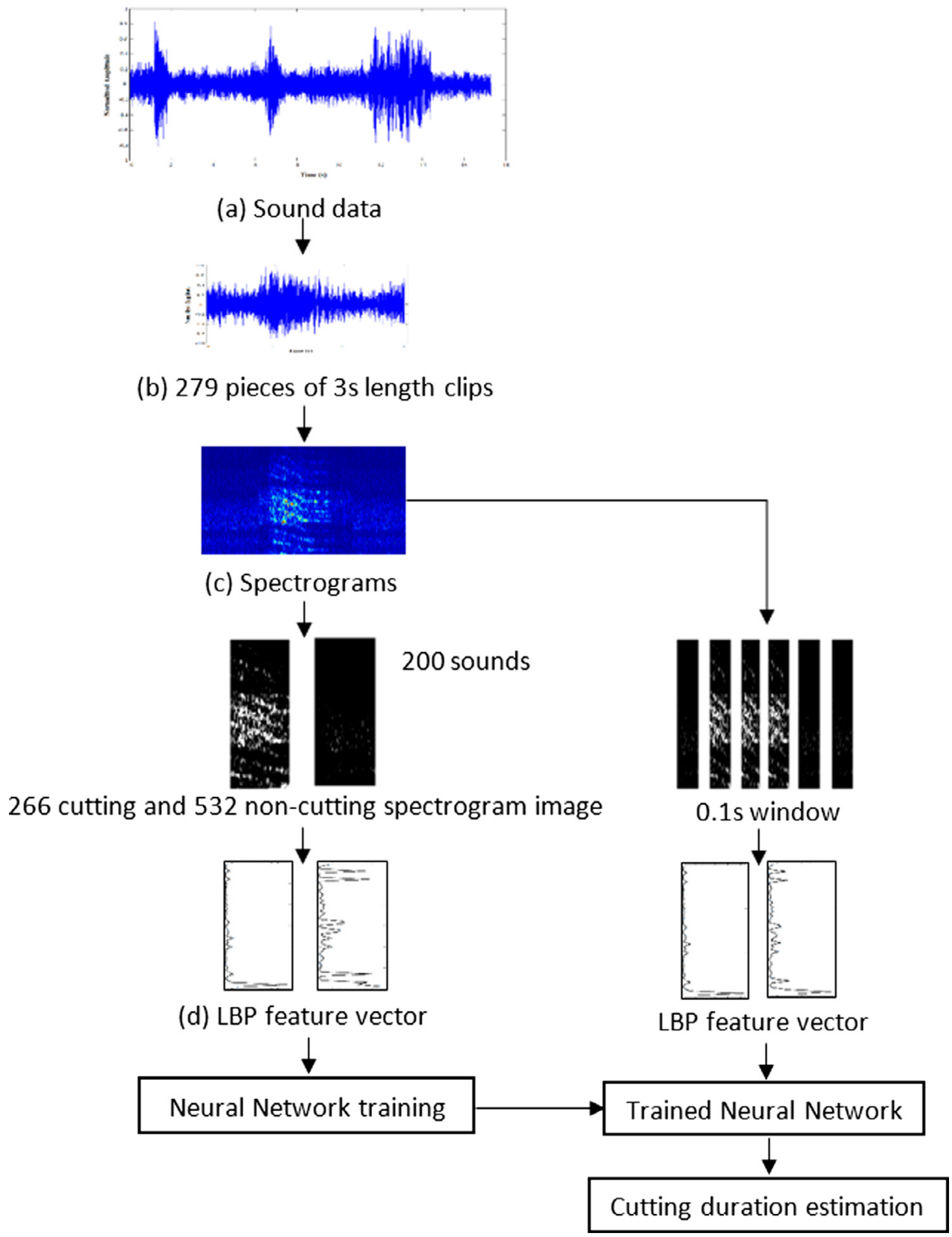

Fig. 3. Block diagram of the proposed feature-extraction and DBH prediction process.

3. Results

3.1. MD and DBH

A linear regression between duration and stem DBH (Fig. 4) for the prediction set data showed a positive, non-zero slope (p-value $<$ 0.001) with R^2 of 0.75. The RMSE of the model was 3.8 cm, or 14% given the mean stem diameter (26.8 cm). The results were similar for the validation dataset, with RMSE of the predictions equal to 4.6 cm, or 16% of their mean.

The major error source of in determining the duration was strong noises at the edges of a cut. Fig. 5 shows about 18 s of a recording that included three tree cuts. Two of the cuts, shown within the black box, were visibly distinguishable because of the relatively large amplitude increase in the time record, and were also audibly different from background sounds by their tone structure. The duration of the third cut, shown in the red box, was less easily distinguished. The amplitude change was still visible but the sound itself was corrupted by additional noise. In that case the saw blade made contact with the ground after severing. Sound interference from a variety of sources was relatively common, in general, which increased the error in determining the duration by human hearing. But for the cuts in this study for which diameters were known, the effect of interference could be accounted for with reasonable assurance of accuracy.

3.2. Automated methods for measuring cut duration (AD)

Although the linear relationship was significant (p-value $<$ 0.0001, R^2 = 0.83), there was a bias between the estimates (non-zero intercept) and the slope was also statistically different from 1 (P $<$ 0.01, Fig. 6). The average difference in duration estimates between the two methods was positive ($\frac{\sum_{i=1}^{200}(AD(i) - MD(i))}{200} = 0.11s,\quad standard\ \ deviation\ \ \sigma = 0.26s$), implying the AD method tended to overestimate the duration relative to a human. The bias between MD and AD estimates was not consistent, however, and AD durations were less than MD for trees having cut durations less than 0.5 s (smaller stems). This result highlighted the

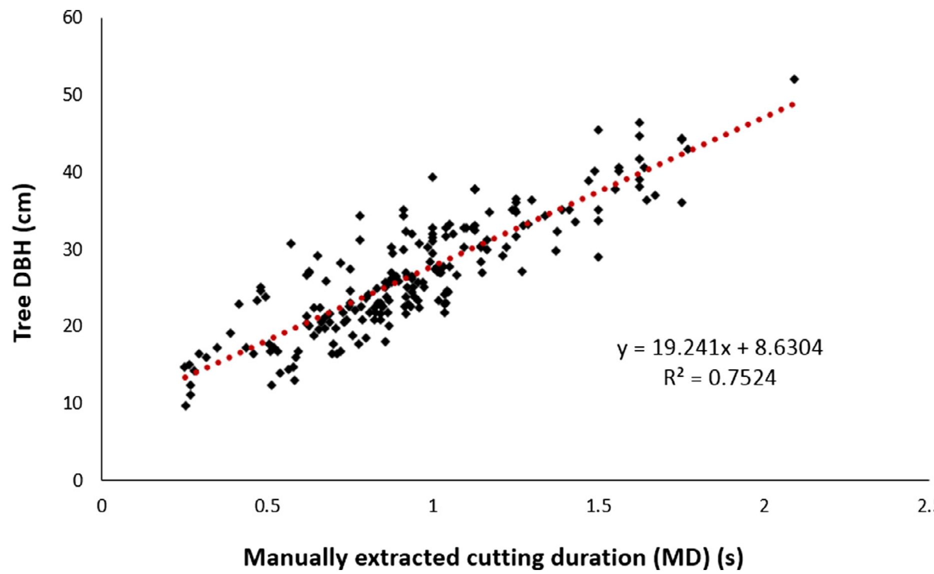

Fig. 4. The linear regression plot of 200 trees' DBH and their cutting duration (MD).

effect of the relatively long window used to estimate spectrograms (about 0.1 s) and the low resolution afforded relative to the duration of the cutting event.

The linear regression between AD and the DBH for the 200 clips in the training data set was significant (p-value < 0.001) and R^2 was slightly better than the MB predictor (0.79 vs 0.75, Fig. 7). The RMSE of calibration based on the neural net estimator in diameter (3.50 cm, 13.1%) was also similar to that from the MD value (3.83 cm, 14.3%). The RMSE of prediction, the AD of which was calculated from the reserved 79 sounds, was 3.57 cm or 12.1%. The result was also better than manual (MD) method (4.6 cm, 15.6%).

The conditions for two sites were different. Including a 'site' covariate in a covariance model for DBH based on the AD metric showed that both site and its interaction with duration, however, were not significant (p-value = 0.57, 0.84, respectively).

4. Discussion

4.1. Method overview

Automatically measuring cut duration from sound was fundamentally a classification problem with the goal of accurately resolving the audio stream into cut/not-cut classes with sufficient time resolution. Among several approaches attempted, such as Mel-frequency cepstral coefficients (MFCCs) with NN (as in (Grézl and Černocky, 2009)) and time-domain duration & features (Rafezi et al., 2012), the LBP duration measurement found to be most successful in this study used spectral characteristics of the audio stream as input features to a neural net model for classification.

Other non-manual approaches to mapping tree sizes can provide much improved accuracy, LiDAR methods being the most obvious.

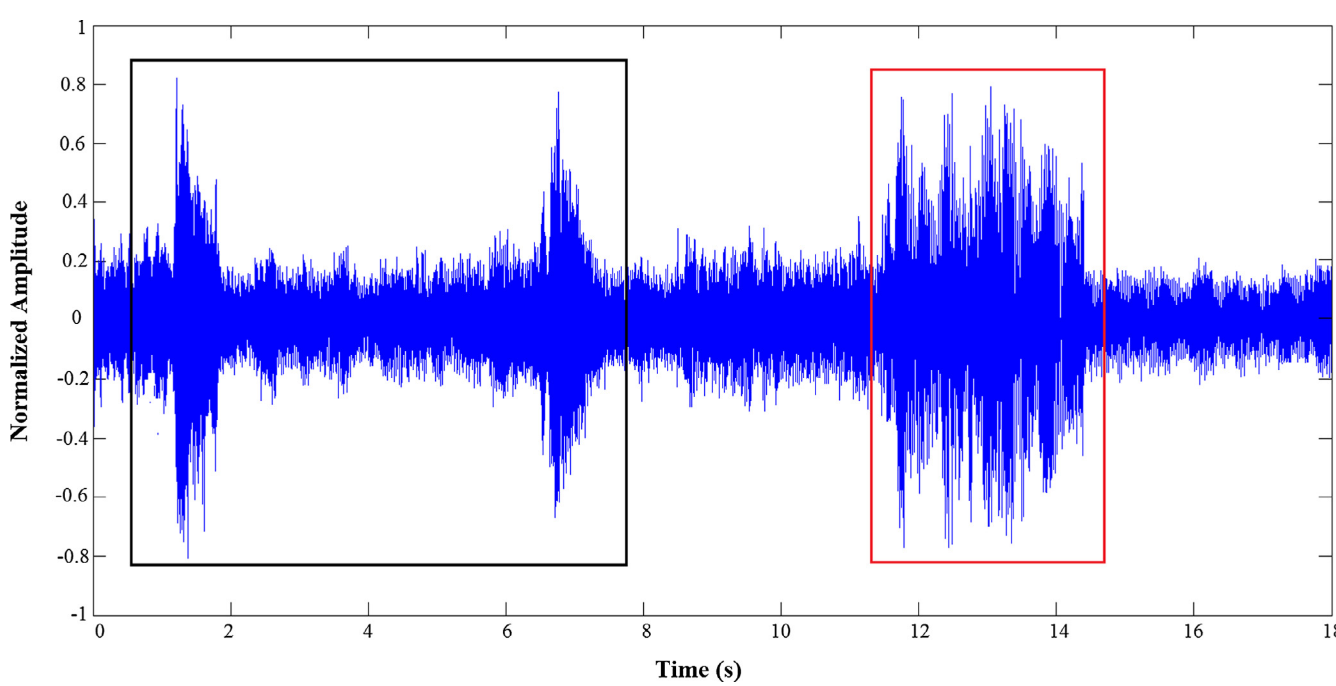

Fig. 5. 18 seconds feller-buncher working sound example. Black box shows two cuts without loud noise influence, red box is a single cut corrupted by loud noise. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

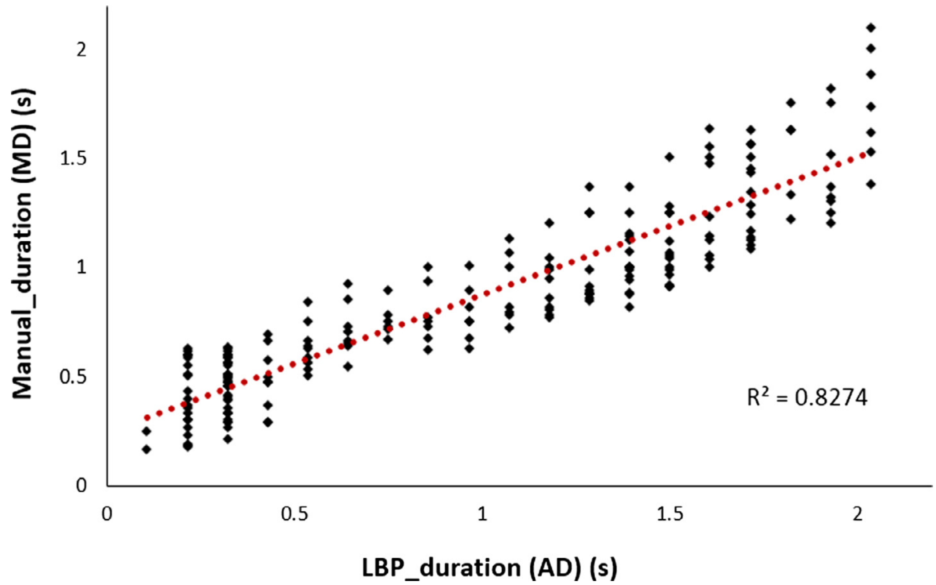

Fig. 6. The linear regression plot of AD and MD (200 sounds in calibration set).

Popescu (2007) found RMSE of individual tree DBH estimates to be 4.9 cm for pines in east Texas based on airborne LiDAR scanning. Murphy et al. (2010) found mean bias of terrestrial laser scanning measurements of base diameter of radiate pine stems to be within 20–25 mm of manual measurements, roughly half the error rate of a harvester on the same stems. Although undoubtedly more accurate, especially the ground-based methods, LiDAR might not have the immediacy of measurements at the time of harvest.

The sound predictions of diameter were less accurate than reported for CTL harvesting heads in similar-sized trees. Scandinavian standards for DBH accuracy require 90% of measurements to be within plus or minus 4 mm of a hand measurement (Andersson and Dyson, 2001). Errors for the proposed method, quantified as the standard error of prediction when compared to a hand-measured DBH, were on average more than 10 times higher than that standard. Using sound duration as a means of driving individual tree value optimization was, therefore, not feasible given accepted accuracy constraints.

As a tool for characterizing stands, however, the system could serve as a rough 'yield mapping' sensor. Numbers of samples to provide a 95% confidence interval on an estimate of mean diameter plus or minus 0.5 cm would be on the order of 325 given the prediction variability observed in this study. That corresponds to about the expected stocking level per hectare for mature pine plantations in the US South. The sensor could be used to accurately predict per hectare inventories of, for example, felled versus processed timber on a tract undergoing harvest. Such information could give mill procurement managers the confidence to implement an inventory control system closer to a just-in-time approach, rather than the more traditional method of stockpiling before periods of expected shortages. This could lower carried inventory costs and allow greater flexibility when purchasing wood in times of weather-related scarcity.

4.2. Error sources and limitations

Both the MD and AD approaches used to estimate cut duration had accuracy limitations. In the case of the manual estimates, the start and end of the cut was a matter of interpretation on the part of the listener and could not be relied upon to be completely consistent. Another

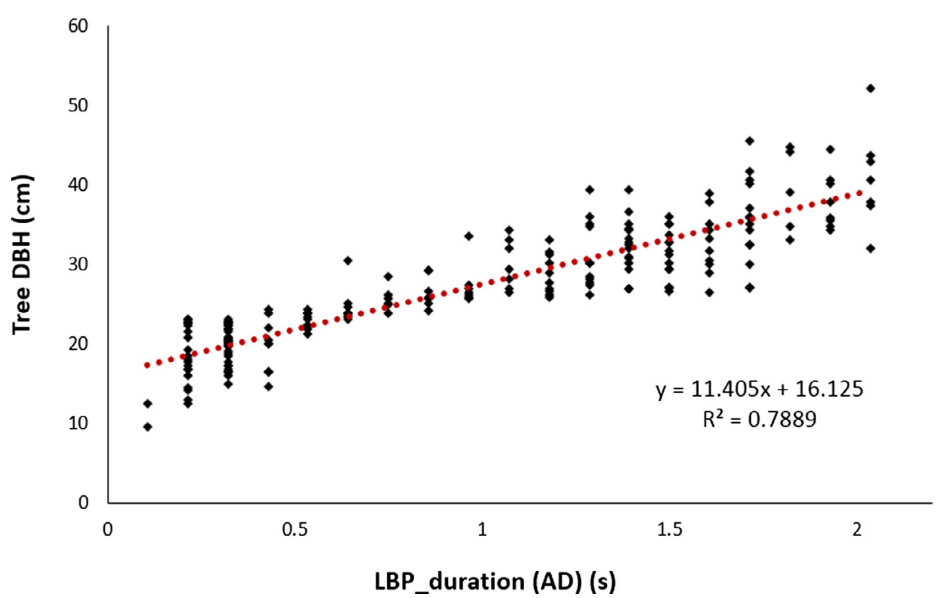

Fig. 7. The linear regression plot of AD and DBH in calibration set (200 sounds).

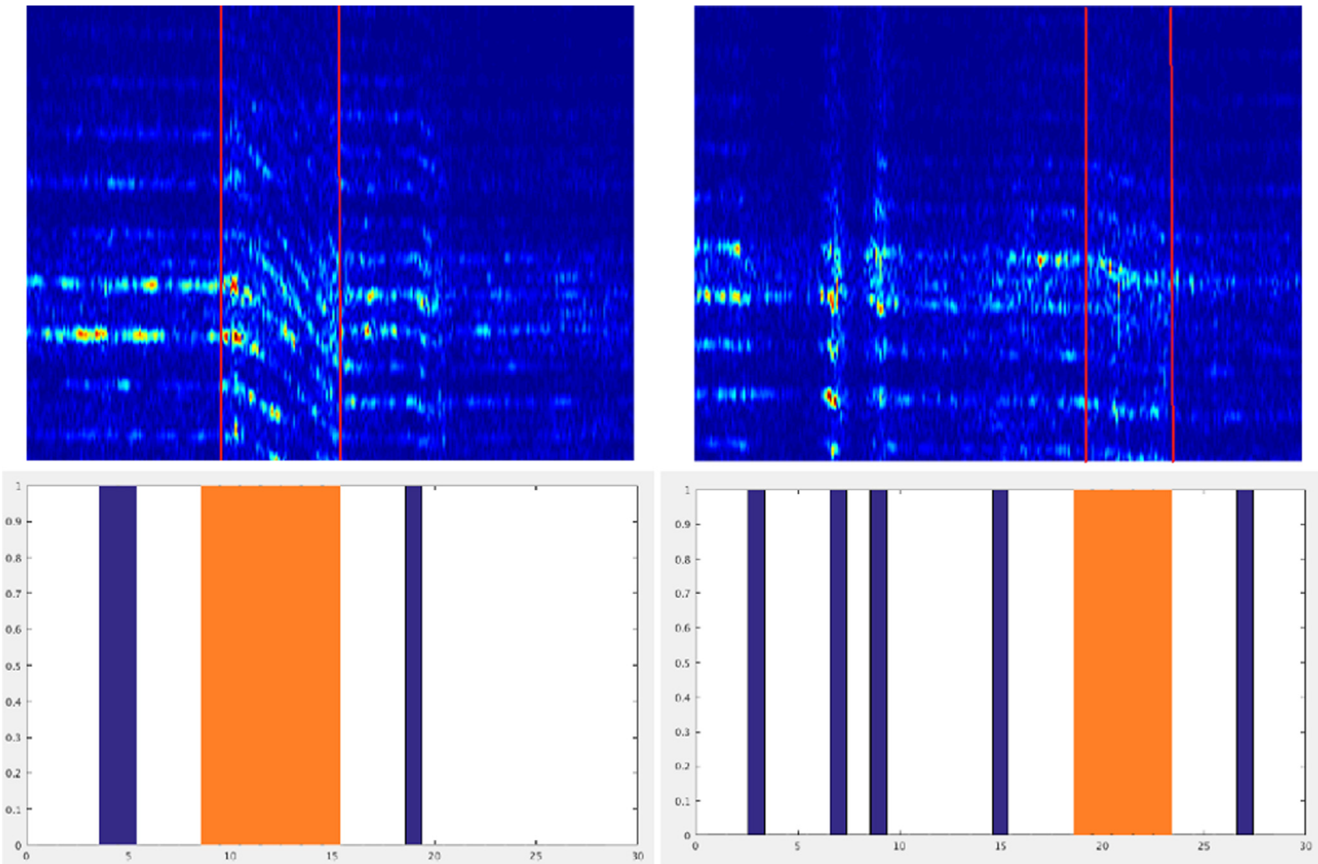

Fig. 8. Spectrogram and NN output examples of cuts corrupted by pseudo-cutting noises. (a) Saw contacted harvested trees during the cut, (b) tree cut inside of thick brushes. Marked area (above) in the spectrogram is MD, and the orange area of classifier (below) is AD. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

limiting factor was noise corruption from sources other than the severing of the stem masking the true start or end of a cut event. For the automated duration estimates, the LBP feature-based neural network classifier could at least be relied upon to be consistent. Other factors, however, affected its performance, particularly the time resolution of the sound features being long relative to the average cut duration. The window of time over which the LBP features were estimated in the classification scheme was 0.1 s, compared to the average cut duration of about 0.9 s. There were many instances of pseudo-cutting sounds identified by the neural net classifier in many of the clips used in this study, mostly from contact of the saw with other brush or small stems while in the process of felling the tree. Fig. 8, for example, shows spectrograms from two cutting events. In both cases there were multiple periods of time identified as being of class 'cut', but selecting the longest contiguous period as the cut duration eliminated the pseudo-cutting events from consideration. If, however, the pseudo-cut sound happened to occur near the start and/or end of the actual cut, the duration could easily be off by up to 20%. Although the classifier was robust in excluding pseudo-cutting sounds other than the actual stem severing, it was still potentially subject to large errors due to noise.

Although the AD method developed for this study worked as well as a human in estimating cut duration, its use in a real-time system for implementing a yield-mapping scheme would require significant development. The largest hurdle to overcome would be distinguishing cut events of non-crop trees from marketable stems. The 3-s clips used in the study excluded a great deal of the noise events present in the extended recordings. The saw constantly contacted other stems and brush while the feller-buncher was operating in the stand and those events would likely be classified as cuts by the neural net developed for this work. Although their short duration would probably eliminate many of them from consideration as true cuts, there would likely be situations encountered when it would not and additional information would be necessary to effectively classify sounds. Some of that additional information could be derived from operator inputs to the machine, and another source might be some of the tonal changes obvious to a human listener, but not really exploited in the neural net classification scheme. For the cut examples seen in Fig. 8, there was a visually distinct drop in frequency present while severing a crop stem. Incorporating these spectral data into the classifier using additional features would probably be necessary in a production system.

Data for this study were collected from two similar fellers operating on two distinct sites. The machines were of the same brand but different ages, both with experienced operators, and stands were both plantation pine of different ages and conditions, but the effects of these variations were confounded in the results. There were some variations in observed response, however, that suggested differences in DBH prediction between the two sets of conditions. Although the site factor was not significant in the model, there was an observable reduction in variability between the two tests. The average difference and standard deviation between MD and AD estimates of cut duration was 0.11 s and 0.26 s, respectively, when all observations were included. If limited to the second site, the values were smaller (mean difference = 0.04 s and the standard deviation = 0.12 s). Although the effects were confounded with machine (and other) factors and there was no authoritative way of separating them, the second site was observed to have many fewer understory stems. The lower understory stem density likely decreased the probability of noise interference in the sound data and improved accuracy and precision of the automated cut duration estimates.

4.3. Potential applications

Based on these results, the use of cut duration, at least a duration derived solely from sound, to make value decisions in tree-length harvest of individual stems was of limited practical value. Estimation of mean size for a group of trees, however, could be valid provided there were enough samples to satisfy any confidence interval constraints imposed on mean diameters. In that application, the sound-based size estimator could be used as a tool for rough yield mapping and would have potential value as a tool for forest management and procurement decision-making. Measurement of stem size in either cut-to-length or tree-length harvesting is a difficult process, but complicated in tree-length felling because the cut stems are not handled individually. Using sound to predict stem size is attractive in this case because it is non-contact and would not require significant structural modification to the felling head to implement. To be useful, however, the method must be augmented to improve accuracy, perhaps with additional measures. Sound recordings for this work were taken from video data. At the expense of additional processing, the image data could, for example, be used to estimate tree size during cutting.

5. Conclusion

As was expected, the tree size is linearly correlated with its cut duration ($R^2 \geq 0.75$). A method of using LBP features and NN classifier was proposed to automatically extract cut duration from the spectrogram generated from an audio recording of the tree felling using a hotsaw. Duration measurements from the automated system were related to manual estimates, but were biased, probably as a result of low time resolution and pseudo-cutting noise corruption. The LBP_NN method achieved a slightly better result in estimating tree size, compared with manual extracted duration. The Overall, the best results indicated individual stem DBH could be predicted with 90% accuracy within plus or minus approximately 7 cm with either manual or automated duration measurement.

Acknowledgements

This material is based upon work supported by the Department of Energy, Office of Energy Efficiency and Renewable Energy (EERE), under Award Number DE-EE0006639. We would also like to thank Fuller's Logging (Lanett, AL) and Industree Companies (Wetumpka, AL) for providing the opportunities in data collection, thank Dr. Thomas Gallagher from School of Forestry and Wildlife Sciences, Auburn University, AL for measuring tree DBH.

References

Akariman, Q., Jati, A.N., Novianty, A., 2015. Face recognition based on the Android device using LBP algorithm. In: 2015 International Conference on Control, Electronics, Renewable Energy and Communications (ICCEREC), Bandung, pp. 166–170. < https://doi.org/10.1109/ICCEREC.2015.7337037 > .

Amoodeh, M.T., Khoshtaghaza, M.H., Minaei, S., 2006. Acoustic on-line grain moisture meter. Comput. Electron. Agric. 52, 71–78.

Anami, B.S., Pagi, V.B., 2013. Localisation of multiple faults in motorcycles based on the wavelet packet analysis of the produced sounds. IET Intel. Transport Syst. 7 (3), 296–304. https://doi.org/10.1049/iet-its.2013.0037.

Andersson, B., Dyson, P., 2001. Log measuring accuracy of harvesters and processors. In: 2001 Council on Forest Engineering (COFE) Conference Proceedings: "Appalachian Hardwoods: Managing Change", Snowshoe, July 15–18.

Azmy, M.M., 2015. Classification of lung sounds based on linear prediction cepstral coefficients and support vector machine. In: 2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT), Amman, pp. 1–5. < https://doi.org/10.1109/AEECT.2015.7360527 > .

Chang, C.Y., Chang, Y.P., 2013. Application of abnormal sound recognition system for indoor environment. In: 2013 9th International Conference on Information, Communications & Signal Processing, Tainan, pp. 1–5. < https://doi.org/10.1109/ICICS.2013.6782772 > .

Cho, J.W., Park, H.M., 2016. Independent vector analysis followed by HMM-based feature enhancement for robust speech recognition. Signal Process. 120, 200–208. https://doi.org/10.1016/j.sigpro.2015.09.002.

Escolà, A., Planas, S., Rosell, J.R., Pomar, J., Camp, F., Solanelles, F., Gracia, F., Llorens, J., Gil, E., 2011. Performance of an ultrasonic ranging sensor in apple tree canopies. Sensors 11, 2459–2477.

Faudzi, S.A.A.M., Yahya, N., 2014. Evaluation of LBP-based face recognition techniques. In: 5th International Conference on Intelligent and Advanced Systems (ICIAS), pp. 1–6. < https://doi.org/10.1109/ICIAS.2014.6869522 > .

Ferreira, A., Alarcão, D., 2016. Real-time blind source separation system with applications to distant speech recognition. Appl. Acoust. 113, 170–184. https://doi.org/10.1016/j.apacoust.2016.06.024.

González-Hernández, F.R., Sánchez-Fernández, L.P., Suárez-Guerra, S., Sánchez-Pérez, L.A., 2017. Marine mammal sound classification based on a parallel recognition model and octave analysis. Appl. Acoust. 119, 17–28. https://doi.org/10.1016/j.apacoust.2016.11.016.

Grassi, M., 2005. Do we hear size or sound? Balls dropped on plates. Percept. Psychophys. 67 (2), 274–284.

Grézl, F., Černocky, J., 2009. Audio surveillance through known event classification. Radioengineering 18 (4), 671–675.

Grozdić, Đ.T., Jovičić, S.T., Subotić, M., 2017. Whispered speech recognition using deep denoising autoencoder. Eng. Appl. Artif. Int. 59, 15–22. https://doi.org/10.1016/j.engappai.2016.12.012.

Heikkilä, M., Pietikäinen, M., Schmid, C., 2009. Description of interest regions with local binary patterns. Pattern Recogn. 42 (3), 425–436. https://doi.org/10.1016/j.patcog.2008.08.014.

Heng, C.K., Yokomitsu, S., Matsumoto, Y., Tamura, H., 2012. Shrink boost for selecting multi-LBP histogram features in object detection. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition, pp. 3250–3257. < https://doi.org/10.1109/CVPR.2012.6248061 > .

Huang, H.B., Li, R.X., Yang, M.L., Lim, T.C., Ding, W.P., 2017. Evaluation of vehicle interior sound quality using a continuous restricted Boltzmann machine-based DBN. Mech. Syst. Sig. Process. 84 (A), 245–267. https://doi.org/10.1016/j.ymssp.2016.07.014.

Kan, J., Li, W., Sun, R., 2008. Automatic measurement of trunk and branch diameter of standing trees based on computer vision. In: 2008 3rd IEEE Conference on Industrial Electronics and Applications, Singapore, pp. 995–998. < https://doi.org/10.1109/ICIEA.2008.4582664 > .

Kojima, T., Ijiri, T., White, J., Kataoka, H., Hirabayashi, A., 2016. CogKnife: Food recognition from their cutting sounds. In: 2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Seattle, WA, pp. 1–6. < https://doi.org/10.1109/ICMEW.2016.7574741 > .

Korucu, M.K., Kaplan, Ö., Büyük, O., Güllü, M.K., 2016. An investigation of the usability of sound recognition for source separation of packaging wastes in reverse vending machines. Waste Manage. 56, 46–52. https://doi.org/10.1016/j.wasman.2016.06.030.

Leitner, T., Stampfer, K., Visser, R., 2014. Analysing log length measurement accuracy of harvester and processor heads. Austrian J. For. Sci. 129–146.

Li, Y., Wu, Z., 2015. Animal sound recognition based on double feature of spectrogram in real environment. In: 2015 International Conference on Wireless Communications & Signal Processing (WCSP), Nanjing, pp. 1–5. < https://doi.org/10.1109/WCSP.2015.7341003 > .

Lin, M., Zhong, S., Lin, L., 2015. Chicken sound recognition using anti-noise mel frequency cepstral coefficients. In: 2015 Third International Conference on Robot, Vision and Signal Processing (RVSP), Kaohsiung, pp. 224–227. < https://doi.org/10.1109/RVSP.2015.60 > .

Mayorga, P., Valdez, J., Druzgalski, C., Zeljkovic, V., 2016. Heart and lung sounds based events classification. In: 2016 Global Medical Engineering Physics Exchanges/Pan American Health Care Exchanges (GMEPE/PAHCE), Madrid, pp. 1–6. < https://doi.org/10.1109/GMEPE-PAHCE.2016.7504614 > .

McDonald, T.P., Rummer, R.B., Grift, T.E., 2003. Diameter sensors for tree-length harvesting systems. In: 2nd Forest Engineering Conference, May 12–15, Växjö, Sweden, pp. 45–54.

Murphy, G.E., Acuna, M.A., Dumbrell, I., 2010. Tree value and log product yield determination in radiate pine (*Pinus radiata*) plantations in Australia: comparisons of terrestrial laser scanning with a forest inventory system and manual measurements. Can. J. For. Res. 40 (11), 2223–2233. https://doi.org/10.1139/X10-171.

Ojala, T., Pietikäinen, M., Mäenpää, T., 2002. Multi-resolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. 24 (7), 971–987.

Olivera, A., Visser, R., Acuna, M., Morgenroth, J., 2016. Automatic GNSS-enabled harvester data collection as a tool to evaluate factors affecting harvester productivity in a Eucalyptus spp. harvesting operation in Uruguay. Int. J. Forest Eng. 27 (1), 15–28. https://doi.org/10.1080/14942119.2015.1099775.

Owens, A., Isola, P., McDermott, J., Torralba, A., Adelson, E.H., Freeman, W.T., 2016. Visually indicated sounds. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2405–2413.

Pagi, V.B., Wadawadagi, R.S., Anami, B.S., 2015. An acoustic approach for multiple fault diagnosis in motorcycles. In: 2015 International Conference on Soft-Computing and Networks Security (ICSNS), Coimbatore, pp. 1–7. < https://doi.org/10.1109/ICSNS.2015.7292413 > .

Peng, J., Wang, C., Jiang, P., Lau, S.K., 2016. Investigation of Chinese word recognition scores of children in primary school classroom with different speech sound pressure levels. Appl. Acoust. 110, 235–240. https://doi.org/10.1016/j.apacoust.2016.03.026.

Popescu, S.C., 2007. Estimating biomass of individual pine trees using airborne lidar. Biomass Bioenergy 31 (9), 646–655. https://doi.org/10.1016/j.biombioe.2007.06.022.

Rafezi, H., Behzad, M., Akbari, J., 2012. Time domain and frequency spectrum analysis of sound signal for drill wear detection. Int. J. Comput. Electr. Eng. 4 (5), 722–725.

Satpathy, A., Jiang, X., Eng, H.L., 2014. LBP-Based edge-texture features for object recognition. IEEE Trans. Image Process. 23 (5), 1953–1964. https://doi.org/10.1109/TIP.2014.2310123.

Skogforsk, 2015. The Forestry Research Institute of Sweden, StanForD. Retrieved from: < https://www.skogforsk.se/english/projects/stanford > (accessed 08 October 2018).

Thamrin, N.M., Arshad, N.H.M., Adnan, R., Sam, R., Razak, N.A., Misnan, M.F., Mahmud, S.F., 2013. Tree diameter measurement using single infrared sensor for non-stationary vehicle context in agriculture field. IEEE 4th control and system graduate research colloquium (ICSGRC) 38–42.

Valenti, M., Tonelli, D., Vesperini, F., Principi, E., Squartini, S., 2017. A neural network approach for sound event detection in real life audio. European Signal Processing Conference (EUSIPCO). < https://doi.org/10.23919/EUSIPCO.2017.8081712 > .

Vandermeulen, J., Bahr, C., Johnston, D., Earley, B., Tullo, E., Fontana, I., Guarino, M., Exadaktylos, V., Berckmans, D., 2016. Early recognition of bovine respiratory disease in calves using automated continuous monitoring of cough sounds. Comput. Electron. Agric. 129, 15–26. https://doi.org/10.1016/j.compag.2016.07.014.

Yogesh, C.K., Hariharan, M., Yuvaraj, R., Ruzelita, N., Adom, A.H., Sazali, Y., Kemal, P., 2017. Bispectral features and mean shift clustering for stress and emotion recognition from natural speech. Comput. Electr. Eng. 62, 676–691. https://doi.org/10.1016/j.compeleceng.2017.01.024.

Zheng, Y., Guo, X., 2017. Identification of chronic heart failure using linear and nonlinear analysis of heart sound. 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Seogwipo, pp. 4586–4589. < https://doi.org/10.1109/EMBC.2017.8037877 > .