Published in final edited form as:

Int J Monit Surveill Technol Res. 2014 July 1; 2(3): 31–49. doi:10.4018/IJMSTR.2014070102.

Vehicle Engine Classification Using Spectral Tone-Pitch Vibration Indexing and Neural Network*

Jie Wei¹, Karmon Vongsy², Olga Mendoza-Schrock², and Chi-Him Liu¹

Jie Wei: wei@cs.ccny.cuny.edu; Karmon Vongsy: karmon.vongsy@us.af.mil; Olga Mendoza-Schrock: mendoza.7@wright.edu; Chi-Him Liu: cliu06@citymail.cuny.edu

¹Department of Computer Science, City College of New York, USA

²Air Force Research Laboratory, Wright-Patterson AFB, USA

Abstract

As a non-invasive and remote sensor, the Laser Doppler Vibrometer (LDV) has found a broad spectrum of applications in various areas such as civil engineering, biomedical engineering, and even security and restoration within art museums. LDV is an ideal sensor to detect threats earlier and provide better protection to society, which is of utmost importance to military and law enforcement institutions. However, the use of LDV in situational surveillance, in particular vehicle classification, is still in its infancy due to the lack of systematic investigations on its behavioral properties. In this work, as a result of the pilot project initiated by Air Force Research Laboratory, the innate features of LDV data from many vehicles are examined, beginning with an investigation of feature differences compared to human speech signals. A spectral tone-pitch vibration indexing scheme is developed to capture the engine's periodic vibrations and the associated fundamental frequencies over the vehicles' surface. A two-layer feed-forward neural network with 20 intermediate neurons is employed to classify vehicles' engines based on their spectral tone-pitch indices. The classification results using the proposed approach over the complete LDV dataset collected by the project are exceedingly encouraging; consistently higher than 96% accuracies are attained for all four types of engines collected from this project.

Keywords

Remote Sensing; Fourier transforms; Indexing; Band-pass filters; Multi-layer neural network; Classification

INTRODUCTION

The use of a Laser Doppler Vibrometer (LDV) as one important multi-modal imaging instrument has been consistently growing in recent years due to the unique advantage it can provide. An LDV sends out a laser beam to a reflective surface S which is then reflected back to the LDV, the amplitude and frequency of S 's vibrations are extracted from the

*This work was supported in part by the U.S. Air Force Summer Faculty Fellowship Program 2014 under Contract No. FA9550-14-C-0013. This paper is approved for release by AFRL with PA No. 88ABW-2014-4870.

Doppler shift of the reflected laser beam frequency due to the S 's vibrating motion. LDV sensors provide many advantages:

1. Non-contact and non-destructive measurements: No mass or pressure is applied during LDV's measurement process; they are almost entirely harmless (only some high power types may cause damage to human eyes if viewed directly). LDV exerts no additional pain in non-invasive medical applications such as body temperature and pulse monitoring during medical operations, and causes no extra damages in non-intrusive civil engineering applications such as inspections of bridges, railways and buildings (Willemann, Castellini, Revel, & Tomasini, 2004) (Kubota, 2007) and delicate inspections of murals and antique fresco paintings in museums where LDV is the only viable means for inspection purposes (Castellini, Paone, & Tomasini, 1996). By contrast, the small dose of radiation from CT or X-ray in medical applications can have side effects to human cells; and the water penetration and/or corrosion are troublesome side effects in traditional ultrasonic tests.
2. High spatial and spectral resolution: the expansive and wide range of amplitudes and frequencies provided by LDV sensors confer researchers and developers with valuable information in both spatial and frequency domains for ensuing intensive analysis and classification.

The use of LDV in many research and development subjects has become increasingly more popular due precisely to the foregoing advantages. As a non-invasive and remote sensor with the afore-mentioned benefits, LDV measured data are an ideal modality/phenomenology to detect potential threats earlier and provides better protection to society, which is of utmost importance to military and law enforcement institutions. The use of LDV in vehicle classification using signature analysis is of crucial interest to military and law enforcement institutions, but is still in its infancy due to the lack of systematic investigations on its behavioral properties.

Many previous efforts have been made to address the issues of vehicle classification using LDV data with varying degrees of success. In (Wang, Zhu, & Taylor, 2013), Wang and colleagues developed a method to detect and classify civilian vehicles based on multimodal audio-visual features, including visual tokens using global geometric features (aspect ratio profiles) and local structure features (HOGs), as well as various LDV features (MFCCs, short term energy, etc.). Radial-based support vector machine was utilized for classification purposes. Promising performances have been reported thereof. In (Smith, Mendoza-Schrock, Kangas, Derking, & Shaw, 2014) a hierarchical vehicle classification approached using laser-vibrometry data was developed, where a wide array of time and frequency domain features such as spectral flux, MFCC, and number of zero-crossings are tested and automatically selected using a parameter selection procedure to generate a decision tree of different types of vehicles such as vans and sedans. In (Sigmund, Shelley, Bauer, & Heitkamp, 2012) the auto-correlation function of LDV signals was employed as the workhorse to distinguish engine type, speed, and number of cylinders with impressive precision. Averbuch et al. developed a diffusion map based framework to detect moving vehicles based on wavelet packets within the dynamic programming framework (Averbuch & Hochman, 2010). In (Ma, et al., 2013), a prototype automatic vehicle classification system

was fabricated, where a grid of accelerometers are installed on roadways to characterize road vibrations and the number of axles is classified accordingly. The corresponding performance compared with the ground truth is exceedingly valuable at about 99%. Note here accelerometer data is used instead of LDV data in the classification. The relevance to our LDV sensory data based classification will be revealed later. In (Rashed, 2014) the original MFCC after a PCA dimension reduction was fed to a forward neural network to classify different Arabic speakers. In (Petrovich, Snorrason, & Stevens, 2002), vehicle operating conditions were classified using 11 extracted features, including MFCC and others such as zero-crossings, dominant frequencies and Flux, which was also employed in (Smith, Mendoza-Schrock, Kangas, Derking, & Shaw, 2014).

In this work, in view of the lack of reliable first-hand LDV data for vehicle classification research and development purposes, a pilot project was initiated by the Air Force Research Laboratory (AFRL) in the summer of 2014. The team (led by the second and third authors) consists of AFRL researchers, college professors and students, and private company engineers—all authors are team members. The collected data includes multiple modalities amidst vehicles of different types over five weeks using several different sensors, LDV, and accelerometers with two specific modalities. This paper concentrates on the LDV data.

Based on the collected LDV data and the ensuing intensive analysis, the innate features of LDV data from multiple vehicles are first studied, in particular their main differences from human speech signals. A spectral tone-pitch vibration indexing scheme is next developed to capture the engine's periodic vibrations and the associated fundamental frequencies on the vehicles' surfaces. After extensively administering and exploring more than 10 different well-established classifiers, a 2-layer neural network with 20 hidden neurons is able to deliver the optimal performance to classify vehicles' engines based on the spectral tone-pitch indices in both the cross-validation and the intensive test method. The classification results using the proposed approach over the complete LDV dataset are exceedingly encouraging as consistently higher than 96% accuracies are attained for all four types of vehicle engines collected in this project.

The remainder of this paper is organized as follows. The LDV data used for vehicle classification is presented in the next section; the spectral tone-pitch vibration indexing scheme to capture the spectral and possibly fundamental frequencies on the vehicle's surfaces is detailed in the scheme section. The classification algorithm used in our vehicle engine categorization work is described in the classification section. The experimental results using the LDV data collected by our team is reported in the results section. The paper concludes with final remarks on this work and related research to be conducted in the near future.

LDV DATA FOR VEHICLE ENGINE CLASSIFICATION

A. Properties and challenges of LDV data

Similar to sound data, LDV data is one-dimensional and interestingly it can be heard using existing sound playing software. Given that LDV data is of sound variety and with the great success in speaker recognition (Kinnunen & Li, 2010) and spoken language processing

(Huang, Acero, & Hon, 2001), much research and development of LDV data analysis follow the track of analogous signal processing. For instance, almost by default, all techniques developed in speaker/speech recognition are also employed in LDV data processing and analysis to date.

Features in the time and frequency domains are extracted using mathematical transformations, such as number of zero-crossings, short-window energy, spectral flux, linear prediction coding (LPC), short-term Fourier Transform (STFT), Mel-frequency cepstrum coefficients (MFCC) (Sigmund, Shelley, Bauer, & Heitkamp, 2012), Hidden Markov Model (HMM), and/or dynamic time warping (DTW) (Bellman, 2003), etc. However, the classification performance reported is poor. In most papers LDV data was almost always fused with other supplementary modality data, such as visual and range features, in order to deliver acceptable classification results, e.g., (Wang, Zhu, & Taylor, 2013). It remains inconclusive if LDV data can serve as another data modality alone, mainly because of the relatively poor performance by current state-of-the-art LDV based classification methods.

One reason why the use of LDV data for vehicle classification purposes is not as successful as speaker and speech recognition widely used in the commercial world lies in the fact that LDV data differs from ordinary human speech signals significantly. The *hearing threshold* in human auditory systems (ear and brain) is ordinarily set at least at 20 Hz (Huang, Acero, & Hon, 2001), that is, the spectral bands lower than this threshold are discarded in the encoding phase, which is one of the enabling technique behind the power of the MP3 compression standard (Solomon, 2007). Conversely, generally lower frequency vibrations, i.e., infrasonic frequencies (< 20 Hz), (Kinsler, Frey, Coppens, & Sanders, 2000) are relevant for vehicle classification: the basic building unit of human speeches, one component of a phoneme, is generally measured in tens of millisecond (ms). This is the reason the default time window for the MFCC is often-times set at around 40 ms (Rabiner & Schafer, 2010). However, vehicles' engines have RPM (revolutions per minute) ranging as low as two to three hundred which is only slightly larger than three Hz. Therefore to sufficiently encode the spectra of vehicles' engines, a time window far longer than 40 ms is needed. Based on our intensive test runs and to provide sufficient information and resolution in the spectral domain, the duration of the basic coding unit should be slightly longer than one second, and in fact the best classification results are attained by setting this duration at 1.25 seconds after intensive experiments. Human auditory systems are studied and exploited for the benefit of effective speech recognition. Several crucial concepts lie behind the immense success of human speaker/speech recognition [1]:

- a. Decibel scale (dB): human auditory systems are not linearly sensitive to the precise magnitudes of a sound, conversely, the psychological sensation curve of human auditory systems, denoted by *loudness*, are rather *compressive*. Thereupon, based on decades of physiological and psychological studies (Hartmann, 2013), the following proportional power law equation holds true to relate the loudness ψ with the sound intensity:

$$\psi \propto I^{0.3}, \quad (1)$$

Where I is the sound intensity values with unit Watts/m². Consequently, the logarithm transform is applied to the sound intensity values to render them better representatives to human auditory systems:

$$L=10 \log(I/I_0), \quad (2)$$

Where I is the intensity of the sound, I_0 is the hearing threshold for human ears that are set at 10⁻¹² Watts/m² by current standard.

- b.** Octave-band collecting: based on the experimental fact that human auditory systems are more sensitive to low frequency amplitudes while less sensitive to differences in increasingly higher frequency bands, which is referred to as *commanding* in acoustics literature (Solomon, 2007). To take advantage of this phenomenon, during the quantization procedure of lossy compression and representation of audio/music signals in forming vectors by using the binning or histogramming operator in the frequency domain, the bin size is doubled as the frequency value increases, thus called octave-band collecting or binning.

One of the main reasons behind the outstanding efficacy of human speaker and speech recognition is the transforms such as the dB and octave-band collecting to convert the original acoustic signals to vectors meaningful to human auditory systems (Lerch, 2012). MFCC takes advantage of the foregoing special properties of human auditory system and is one seminal feature that has achieved great success in many applications of speaker and speech recognition systems. These methods work similar to the effective JPEG and MPEG standards in their respective processing of still images and motion videos, whereof the physiology and even psychology of human eyes and human brain's perceptive powers are studied and exploited. Consequently the visual information is represented in the most economical ways meaningful to human visual systems for successful processing (Szeliski, 2011) (Gonzalez, Woods, & Eddins, 2009).

However, in vehicle classification applications, human auditory techniques are inappropriate or irrelevant. Instead, the vibration data collected by LDV sensors should be treated as a sequence of physical data or time series as the running vehicle engine is the periodic vibrating source propagating/dissipating its energy over the rigid surface of the vehicle as vibrational waves (Kinsler, Frey, Coppens, & Sanders, 2000). The LDV sensor records the vibrations or waves on the surface. The laser beam is extremely small, in terms of micrometers, so the surface where the LDV collects the vibrations can be viewed as a 2-D plane, thus in our ensuing mathematical analysis, although the vehicle surfaces are 3-D surfaces, the geometry around the laser points is in essence two dimensional. The focal points can then assume the phenomenology of periodic vibrations on rigid surfaces and the corresponding mathematics, e.g., partial differential equations (PDE) for waves, and physics, e.g., fundamental frequencies of sheets, modeling, and analysis. Most of the concepts and

ideas targeted for human auditory systems should be carefully re-examined in this new direction of classification endeavors.

Based on the preceding arguments, using features directly from prior successful efforts of human speaker/speech recognition is problematic. It is then recommended to devise new features by taking into account the special nature of vehicle engines in order to achieve desirable classification results.

B. Necessity of LDV sensory data

Major progress has been made in visual (image and video) analysis and understanding, visual feature descriptors such as SIFT (Scale-Invariant Feature Transform), HOG (Histogram of Oriented Gradients), and BRIEF (Binary Robust Independent Elementary Features) (Szeliski, 2011) have been developed with great representative prowess. Viola-Jones Haar cascade method invented in 2001 (Viola & Jones, 2004) has been widely installed in most digital cameras for human face detection due to its exceedingly high accuracy rates. However, in military applications, namely detecting vehicles of potential threat, visual features have major shortcomings, that is, they are easily deceived by the enemy or criminals intended to hide their identities—they can change or camouflage the visual outlooks of their vehicles to render themselves unidentifiable.

The data collected by LDV are reflective of the engine and surface vibrations of the vehicle and thus extremely hard, if not entirely impossible, to counter-measure. Therefore LDV can better penetrate the authentic identities of vehicles and become extremely useful for military and law enforcement institutions for target detection and Intelligence, Surveillance and Reconnaissance (ISR) purposes that are important to national and international security. This capability promotes LDV sensory data therefore is vehemently explored and examined in this work initiated and supported by the AFRL.

SPECTRAL TONE-PITCH VIBRATION INDEXING SCHEME

A. Choice of the indexing duration s

The first issue in the indexing scheme is the choice of coding duration, s . The infrasonic details are needed to reflect the vehicle engine's vibrations in their normal operating conditions. As stated previously, the basic indexing unit s of the LDV index should be of length significantly longer than speech or speaker encoding. Assuming in the *idle* condition, an engine is in the neighborhood of approximately 200 RPM, which translates to 3.33 Hz, the duration should be at least 0.3 seconds, far larger than the default 40 ms for speech coding, e.g., MFCC. When vehicles perform under different operating conditions, the RPM of the engine may differ by a magnitude smaller than 200. This indicates that to provide rich details of the different operating conditions of a vehicle's engine the duration should be able to reflect the meaningful delta δ in the RPM, namely

$$s=60/\delta, \quad (3)$$

where 60 (seconds/minute) is used to convert the unit from minute in RPM to second. Based on our intensive Monte Carlo studies, we discovered that by setting $s=1.25$ seconds, the

resulting vibration index attained the optimal performance consistently. According to Eq. (3), the corresponding RPM δ is 48, signifying that an RPM difference less than 48 has minimal impact on the resultant classification procedure. This aligns well with our observations that uncertainties such as the engine noises, road conditions, and the driver's indistinct pressure changes make it virtually impossible to keep the RPM constant. If s is set at a much higher value, akin to 3 seconds, then finer grained details can be received. However, it is consistently observed from our empirical studies that the classification performance and the representational power is far worse than when setting $s = 1.25$ seconds. The major reason behind this degenerate result is that more noise is introduced during this longer duration which significantly increases the signal-to-noise ratio in the corresponding longer LDV vectors.

B. What should be indexed

After resolving the duration issue, the next step is what should be indexed. The LDV sensor used in our project measures the velocity v of the vibrations at the measured points. Recall that v is the first derivative of the unknown vibrating function u :

$$v = u_t \quad (4)$$

It was revealed that the vibration surrounding the point whose vibrating movements are collected by the LDV sensor is essentially planar; hence the corresponding 2-D wave PDE is as follows

$$u_{tt} = c^2(u_{xx} + u_{yy}) = c^2 \Delta u, \quad (5)$$

where u_{tt} , u_{xx} and u_{yy} and the partial derivatives for time, x-axis, and y-axis, respectively, is the 2-D Laplacian operator; while c is the transmission speed of the vibrations on the vehicle surface, which is roughly viewed as a constant for all vehicles in this work. From Eqs. (4, 5), to effectively index the LDV data in a way uniquely determined by the engine's vibrations and propagations on the surface, the raw LDV data V is inappropriate. Instead, its time derivative V_t should be employed. V_t is precisely the accelerations measured by the accelerometers, which is empirically checked and confirmed by our accelerometer data collected which was collected simultaneously in this project. Eq. (5) dictates that the resultant values are time-invariant and only depend on u that is caused by the engine's vibration. Thereby the data vector of duration 1.25 seconds is the time derivative of the LDV sensory data.

C. Fourier analysis for tone indexing

Given that the vibrating source during s seconds is the vehicle engine, a periodic vibrating force, the v_t are the stacked acceleration magnitudes due to the vibrations of the surface. Within this short duration, $s = 1.25$ seconds, it is unlikely the engine may have too much considerable changes. Even if change occurs, our indexing scheme and classification are not interested in the information about the time locality. Consequently, the analysis can be conducted entirely in the spectral domain¹, i.e., in the Fourier transform domain, similar to

most acoustic analysis work (Kinsler, Frey, Coppens, & Sanders, 2000). The following transform is performed:

$$F_V = |FFT(V_t)|, \quad (6)$$

where $FFT(o)$ is the Fast Fourier transform, which are complex values with both magnitude A and phase θ in the form of $Ae^{i\theta}$; $|o|$ keeping the magnitude while discarding the phase part, i.e., $|Ae^{i\theta}| = A$. The phase information is ignored because it has no significant role in the periodic vibrations of concern in this classification work. Furthermore, it is impossible to obtain the true phase value for each LDV point.

Figure 1 illustrates the F_V 's for four different engines investigated in this pilot study funded by AFRL. In the lower bands, the peaks caused by the engine and the transmission of vibrations at the point being measured align. However, in the higher bands, many peaks may be caused by certain uncontrollable noises. In addition, even peaks that can be explained by the RPM or multiples of fundamental frequencies fail to precisely align. Hence Fourier magnitudes in these relatively higher bands are easily disturbed by noises, then their original values should be discounted so that even if their values and positions in the spectral domain are imprecise their values—after being discounted—will not affect the resultant classification performance adversely. After empirical experimentation, we found that the logarithmic transform is a good remedy:

$$F_{dV}(x) = |\log(F_V(x))| \text{ for } x \geq h_1, \quad (7)$$

where h_1 is a threshold to start the magnitude suppression transform. Eq. (7) dictates that for bands lower than h_1 $F_{dV} = F_V$ are deemed more reliable than those higher. Based on our intensive grid searching procedure, the threshold h_1 is found to be approximately 43, that is, in the spectral domain corresponding to the RPM of vehicle engines above 2500 RPM, this is likely due to the damping efforts made by the vehicle manufacturer in order to keep strong vibrations at bay (Preumont, 2011) and impedance among the main metal surface of vehicles and other materials (Kim, 2010), e.g., rubber or plastic covers, the vibrations picked up by the LDV are not literally reflective of the RPM of engines and thus have to be discounted.

Likewise bands much higher than h_1 the F_{dV} 's are gravely disturbed by noises and are of little value to encode the different operating conditions of the engine, therefore should be decimated altogether. In extremely low spectral bands, such as one or two Hz, which is available since s is larger than one second, are mostly caused by external sources such as wind or the minor direction/position variations, which is unrelated to the engine and the fundamental modes of the rigid surface. These extremely low bands should also be discarded to ensure the vibration index focuses only on the engine's vibrations and its transmissions across the surface. The following band-pass filter (Oppenheim & Schaffer, 2009) is needed to achieve the above two decimation actions:

¹Out of curiosity, several time-domain features, such as short time energy and number of zero-crossings, widely used in speech coding applications, are also tested without making any performance gains on top of the Fourier analysis reported in this work.

$$\mathbf{t}_V = F_{dv}(h_0:h_2), \quad (8)$$

where $h_0 < h_1 < h_2$. The spectral bands lower than h_0 are discarded as irrelevant to the engine's vibrations, while those higher than h_2 , which is larger than the discounting threshold h_1 , are dropped as high frequency noise. This vector $\mathbf{t}_V$ records the meaningful amplitudes in the spectral bands directly related to the periodic vibration of the engine and its associated transmission on the surface, it is denoted by the *tone* of the LDV vector $\mathbf{v}$, reusing the notation in music coding (Lerch, 2012).

D. Second Fourier transform for pitch indexing

The tone index $\mathbf{t}_V$ captures the dominant spectra corresponding to the periodic vibrating motion directly caused by the engine, such as the peaks of maximal amplitudes shown in Figure 1, especially the first three engine classes displayed (I4, 1- and 2-axle). The surface of a vehicle, as a rigid media, has its own fundamental frequencies or modes (Preumont, 2011) as it transmits the vibrations caused by the engine. In structural engineering (Preumont, 2011), the fundamental mode is the lowest frequency whose multiples repeat themselves in the spectral bands. In most mechanical or civil engineering work the greatest common divisor (gcd) of local peaks in the spectra is sought after as the fundamental frequency, or *pitch* in human speech or music recognition work (Huang, Acero, & Hon, 2001). However, for the vehicle surface case, it's likely that more than one mode may be present for reasons listed below:

1. The shape of the vehicle's surface(s) is extremely complex which mostly have more than one mode.
2. Intentional efforts made by vehicle manufacturers to mitigate or dampen the surface vibrations to comfort drivers and passengers may cause undue stresses on the surface materials that will likely reduce the vehicles' lifetime.
3. The differing materials used in most vehicles, such as plastic and rubber parts, will cause non-ignorable impedance (Kim, 2010), the reflected rigid waves may interfere with existing modes thus making the resultant vibrations picked up by the LDV sensors more sophisticated.
4. The strong rigidity of the metal surface damps/diffuses the sharp peaks of any single mode. Consequently the spectral bands adjacent to the modes are of considerable amplitudes, which is the reason why all peaks are of Gaussian shape, however the Laplacian shape of each peak shown in Figure 1 is because of the operation dictated by Eq. (4) performed on the LDV velocity data then the difference of Gaussian data is a Laplacian.

This multiple-mode phenomenology can be observed from the top three spectral curves in Figure 1. The bottom spectral curve associated with V6 engines has one predominant mode; it overpowers all other modes.

The presence of multiple modes makes it problematic to simply use the band passing/stopping filter, and discounted partial vector $\mathbf{t}_V$ as defined by Eqs. (6–8): the fundamental

modes will make their presence in spectral bands far higher than h_2 , even to infinity. For modes of small magnitudes, e.g., the I4 engine curves in Figure 1, the energy loss is minor and may be acceptable. However, in most cases, as shown in the bottom three cases of Figure 1, more prominent in bottom curve for V6 engines, most of the energy due to the fundamental modes will disappear from the $\mathbf{t}_V$ index. An additional vector should then be formulated to capture the energy corresponding to the (multiple) modes.

Although it is impossible for us to estimate the number of modes on the surface, or modeling their shapes and frequencies, the known properties of these modes lend us an approach to encode them through:

1. *Periodicity*: Each mode by itself is periodic in the spectral domain.
2. *Diffused energy*: The diffused energy at each peak is described by a Laplacian shape.

Based on the foregoing properties, if we conduct a second Fourier transform, the energy due to the multiple modes will be mostly concentrated in the lower bands.

Therefore, the following equation is performed in an effort to capture the modes:

$$\mathbf{p}_V = |FFT(F_V)(h_0:h_1)|, \quad (9)$$

F_V as defined in Eq. (6) is the magnitude of the Fourier transform of V . This low-pass filtered vector $\mathbf{p}_V$ encodes the low frequency energy in the spectral curve. Based on our preceding discussions, it reflects the energy provided by the fundamental frequencies or pitches of the surface, which indicates the *pitch* part of the engine and surfaces. The vibration index for the 1.25-second 1-D LDV vector is formed through stacking $\mathbf{t}_V$ and $\mathbf{p}_V$ together.

E. Tone-pitch vibration indexing algorithm

The *spectral tone-pitch vibration indexing scheme* is summarized in the following algorithm:

Spectral tone-pitch vibration indexing

1. The basic representative unit is the vibration data V with duration s second.
2. Compute the first time derivatives of V .
3. Apply Fourier transform and keep the magnitudes only, F_V is obtained as defined by Eq. (6).
4. Conduct Eqs. (7, 8) to get the spectral tone index $\mathbf{t}_V$.
5. Form the spectral pitch index $\mathbf{p}_V$ according to Eq. (8).
6. The vector $\mathbf{f}_V = [\mathbf{t}_V, \mathbf{p}_V]$ is taken as the spectral tone-pitch vibration index for V of duration s seconds.

This index $\mathbf{f}_V$ is the effective feature reflecting the contents of LDV data V that is effectively exploited by the classification algorithm which presented in the next section. In our work,

via intensive bootstrapping Monte Carlo studies using our collected data, h_0 , h_1 and h_2 are set at 4 Hz, 43 Hz, and 83 Hz, respectively, the index is of dimension 120.

The tone-pitch vibration index differs significantly from the MFCC in the following aspects: 1) the duration, 1.25 seconds, is far longer than the standard MFCC (Huang, Acero, & Hon, 2001), measured in milliseconds. 2) The octave band collecting step is not used. 3) Instead of a logarithm transform conducted by MFCC for all spectral data, this drastic detail compressing method is only performed for those with frequency higher than a threshold, cf. Eq. (7). 4) A second FFT dictated by Eq. (9) is used instead of DCT of the logarithm of all filter-bank energies in order to capture the possible repeating pattern of the spectral information present in LDV data. Existing temporal and spectral features as tabulated in (Petrovich, Snorrason, & Stevens, 2002) and (Smith, Mendoza-Schrock, Kangas, Derking, & Shaw, 2014), including MFCC have been extremely useful and effective in a great array of speaker and speech recognition applications. However, from our intensive tests using LDV data previous combinations and exhaustive feature selection procedures for vehicle engine classification are far from acceptable (with 20%~50% accuracy rates) and for that reason a new feature targeted to capture LDV data is necessitated.

ENGINE CLASSIFICATION ALGORITHMS BASED ON TONE-PITCH INDEX

Among all machine learning techniques, supervised learning is the most studied, and among all supervised learning approaches, classification methodologies are most intensively investigated (Murphy, 2012). There are tens of different classification methods to categorize instances to known labels by virtue of various well-studied schemes with solid mathematical and statistical basis. In this work, to keep the error rates below 10% as conventionally required by AFRL, six different types of classification schemes—ten different classification approaches—are appraised as enumerated:

1. Support Vector Machine (SVM): by searching for the largest possible margin separating different classes in the training data, this method is proved to have one of the best generalization performances (Vapnik, 1998).
2. Naïve Bayes (NB): by assuming the independence of the likelihood probabilities, $p(X|Y)$, of different features X 's, the posterior probability of the class label Y is determined by the simple product of the likelihood according to Bayes theorem. Although an extremely simplistic method, NB is a competitive classification method with many applications in mail separation, speech recognition, and visual object categorization (Murphy, 2012).
3. K nearest neighbor (kNN): the known instances are simply saved; when a test instance T arrives, the majority label of T 's k nearest neighbors is given to T . Despite a simple approach, it is rigorously proven that the error rates delivered by kNN is extremely close to the optimal Bayesian error when $k \geq 3$ (Duda, Hart, & Stork, 2001).
4. Random forest (RF): A set of decision trees constructed using randomly chosen features (Breiman, 2001). The consistently great classification performance by random forest spawns immense interest in *ensemble methodology* (Murphy, 2012)

in the machine learning community. Interestingly, given that the decision trees in a random forest are generally of considerable depths, its performance is also attributed to its deep architecture by deep learning community (Deng & Yu, 2014), hence it was also labeled as a *deep learning method*.

5. Boosting: This is another group of ensemble methods. A committee is constructed by a weighted sum of a set of *weak classifiers* (accuracy rates slightly larger than 50%). AdaBoost (Freund & Schapire, 1997), short for adaptive boosting, is most famous in this group. AdaBoost trains the weak classifiers in a sequence; instances that were misclassified by previous classifiers are favored in a later stage. In this work several applicable variants of AdaBoost provided by the *statistics toolbox* of MATLAB were also tested as possible candidate classifiers for our tone-pitch vibration indexes, they are *LogitBoost (LgBoost)*, *RobustBoost (RbBoost)*, *RUSBoost (RsBoost)* and *LPBoost*.
6. Neural network (NN): As an emulation of animal brains, NN is a system of interconnected neurons; each neuron is essentially a non-linear logistic function whose input is the linearly weighted sum of input signals. It was proved rigorously that a 2-layer NN can represent any mathematical function (Bishop, 2007). The back-propagation procedure can effectively train a NN. Because of its similarity to brains, rich mathematical representativeness and efficient construction, NN has been among the most popularly used methodologies in machine learning. SVM is a special NN. Due to the aforesaid development of deep learning (Deng & Yu, 2014) where the network is deeper than two or three, as most NNs are, the NN is also referred to as a *shallow network*. It should be pointed out that the *deep network* as denoted by deep learning community, such as Restricted Boltzmann Machine (RBM) (Hinton & Salakhutdinov, 2006), Convolutional Neural Network (CNN) and Stacked Auto-Encoder (SAE) (Bengio, 2009), *cannot* be simply learned by increasing the number of layers of the NN because the back-propagation procedure is likely trapped in a local minimum thus yielding no gains in performance. In our work, the classifier version, *patternnet*, of NN provided by the *neural network toolbox* of MATLAB (Bishop, 2007) is employed, which does provide an easy interface to change the depth of the network. From our empirical studies, simply increasing the number of layers from two to four or five to make a deeper network architecture can only degenerate the classifier's performance. The deep network should be formed by using different training procedures that can be applied iteratively for each layer without worrying about the local minimum dilemma encountered by the long chain of partial derivative computations involved in the back-propagation procedure, such as the layer-wise pre-training by Hinton's group (Hinton & Salakhutdinov, 2006) and auto-encoder algorithm by Bengio (Bengio, 2009).

To render the classification results more robust to noises, instead of applying the preceding classifiers to the entire 30-second LDV data segment L , the *majority rule* is employed by going through the following steps:

1. L is sliced into N overlapping frames, denoted by $V_i, i=1, \dots, N$.

2. Compute the tone-pitch index f_i for each V_i through the spectral tone-pitch vibration indexing algorithm described above.
3. Classify each f_i to a label c_i using a classification algorithm.
4. All N classification labels, i.e., $(c_1, c_2, \dots, c_N)$, are aggregated to make a majority vote then the label c with the maximal counts is assigned to L .

The actual space the LDV data are situated is a *manifold* of low dimensionality definitely less than 120, the dimensionality of the spectral tone-pitch vibration index as formulated in the preceding section. It is however impossible to know their authentic dimensionality given the data currently available. More data collections, in addition to this pilot study, are needed to gain deeper insights into the statistical behavioral properties of LDV data and their tone-pitch indexes. A simple application of dimensionality reduction (Tan, Steinbach, & Kumar, 2006) may not help to improve the classification performance. In this study we examined some leading dimensionality reduction techniques such as principal component analysis (PCA) and independent component analysis (ICA) (Hyvarinen, Karhunen, & Oja, 2001) to no avail—the corresponding classification performances are consistently worse therefore this step is omitted in its current form.

The foregoing different classification approaches have their own pros and cons, although some such as SVM, RF and AdaBoost are more popular, NB and kNN have their specific niche. Consequently in this work all of them are tested in our pursuit of the optimal classifier(s) for our LDV tone-pitch indexes.

RESULTS

A. Data collection

From June 30 to August 4, 2014, in the parking lot of Wright State University Nutter Center, the team collected LDV data on 12 different vehicles. Many vehicles were collected more than once on different time/days and some vehicles share the same make and year differing only in serial numbers. Our measuring procedure proceeds as follows: The LDV sensor is maneuvered around the vehicle to take 24 to 31 points; location examples include the front bumper, above the front wheels, the front and back doors of the driver and passenger sides, and the back bumper. Vehicles of the four types of engines used in our tests are illustrated in Figure 2.

For each point, the driver was asked to exert five different operating actions: *idle*, *sweep*, *engage drive*, *constant 2000 RPM*, and *power on the AC*. For each action, the LDV recorded a 30 second time sequence. In order to ensure the data quality, three procedures were performed: 1) At the collection point, a reflective tape is applied on the vehicles surface. 2) The distance from the LDV sensor to the collection point is positioned at approximately ten feet. 3) Noise is required to be at a minimum during the 30 second collection period. In total there are four different types of engines contained in this data set, an I4 sedan, 1-axle truck, 2-axle truck, and V6 sedan.

B. LDV data classification: cross validation and testing results

The diagram showing the steps used in this supervised LDV data learning, validation, and test algorithm is depicted in Figure 3.

In all three phases, as described in the *majority rule*, a 30 second duration of LDV data is first sliced into 120 overlapping frames of 1.25 seconds. The spectral tone-pitch index of each frame is formed as the content descriptor of the 1.25-second LDV data. The classifier training procedure using multiple classifiers is first invoked on the training data set. To fine-tune the parameters, another set of data is kept as cross-validation (CV) data and undergoes the same indexing step to obtain the corresponding feature vectors. When the CV results are adequate by readjusting the parameters, a large amount of sequestered data is tested against the learned classifiers.

During the training and CV steps, one *representative vehicle* measured in one day for each of the four different engines is *randomly* chosen to train the classifiers. Four 30-second LDV points for each representative vehicle are randomly selected based on the *stratified sampling approach*: one close to the engine, e.g., on the front bumper or above the wheel, two on the sides, and one on the back bumper, are used as the training data during the training phase as shown in Figure 3. All other points (ranging from 20 to 27) except those points used in the training step are employed as CV data to fine tune the parameters for each classifier.

As can be seen from the preceding CV description, the CV procedure we used differs from the ordinary approach where the size of CV data is less than the training data, e.g., an N-fold CV procedure uses N-1 folds as training data while the other *one* fold as the CV data; the leave-one-out CV procedure uses only one instance as CV data. The CV procedure we used reversed the size designation and thus providing far more data points to serve the role of CV purpose as opposed to training. This is because if the general rule of training CV data partitioning is followed then our LDV data on the same vehicle is overly optimistic leading to almost perfect classification results (close to 100%), and produces minimal control over the parameter retuning which is the primary goal of the CV step. From our *reversed* CV data designation, the validation is more severe than the ordinary N-fold cross validation process. This leads to avoiding the undesirable over-fitting problem and allows for more meaningful results to better retune the parameters in the CV step. In total there are 78 LDV points in the training dataset. The number of CV data points is 378. The optimal parameters are tuned for all selected classifiers in the CV step.

Using the best possible trained and fine-tuned classifiers, the test step classifies all LDV data points other than those used in the training and CV steps amounting to 1949 points. Confusion matrices and Receiver Operating Characteristic (ROC) curves are generated from the known truth labels of the test data compared to the classifier's predicted labels to produce classification quality measures.

As seen in Table 1, in the CV step, while NB failed miserably with the worst accuracy rate of 48.1%, and SVM delivered sub-par performance of 74.9%, the other eight classifiers appraised in our studies achieved impressive performance with six classifiers passing the acceptance bar of 90%, two boosting variants RobustBoost and RusBoost missed the 90%

by a small margin of 89.7% and 88.9%, respectively. The kNN classifier ($k=5$, under 1-norm learned from our data) has the second best performance at 96.6% while the RF and AdaBoost have decent accuracy rates of 95.5% and 95.8%, respectively. The shallow 2-layer NN with 20 hidden neurons achieved 97.4% marking the highest overall performance.

In the first row of Table 2, the confusion matrices for some of the leading classifiers, namely, kNN, RF, AdaBoost and 2-layer NN, are depicted. It can be observed that although kNN has an impressive overall accuracy rate, its accuracy rate for 1-axle engines falls below the 90% threshold. Whereas the best 2-layer NN remains on top for individual accuracy rates with the lowest accuracy from the I4 at 94%.

However, these results are overly optimistic since the CV data and training data are taken from the same vehicle at very short time differences, differing at most by 2 to 3 hours in the collection process. Consequently the performances against the test data that are collected in entirely different times and days with possibly different make and/or serial numbers are more indicative of the true power of the classifiers.

The overall accuracy rates of the ten classifiers for the test data are tabulated in the test row of Table I. Except that the 2-layer NN with 20 hidden neurons has an accuracy rate of 96.1%, all other classifiers no longer exhibit performance at the 90% acceptance bar. This table shows the worst result being the SVM with a mere 39.0%—although still statistically significantly better than a random guess of 25% for 1949 test points; while NB has the second worst accuracy of 60.1%, interestingly its accuracy rate is better than that displayed in the CV step. The kNN classifier cannot keep its impressive standing typical in the CV step as its rate is 67.7%, the overall third worst. The RF classifier and all five boosting variants are clustered ranging from 70.3% to 75.3%, and although these are often-times acceptable results, military applications require higher rates at a 90% acceptance bar. The confusion matrices for the same four classifiers are presented in the second row of Figure 3. Not only is the overall accuracy of the 2-layer NN above 96%, the individual accuracies are no lower than 96% as well. The receiver operating characteristic (ROC) curves for the 2-layer NN in the CV and test steps are illustrated in Figure 4. It can be observed that the area under the curves (AUCs) are significantly larger than 90%, a threshold described for a “good” classifier.

C. Classification results by deep nets

At present the deep network has been the champion in most visual or speech recognition contests (Deng & Yu, 2014). During our parameter grid search, the patternnet with more than one hidden layer was attempted as well, but exhibited no gain. The best performance with two hidden layers is about 5 percent worse than the shallow patternnet reported in Table 2. This is predictable because the back-propagation procedure to learn the NN is likely trapped in a local minimum as the depth of the network increases. To artificially make the shallow network deep using the shallow network training algorithm, the back-propagation procedure is destined to fail (Bengio, 2009). Based on the publicly available deep learning codes provided by Hinton’s group and Palm (Palm, 2012), we tried applicable deep learning methods SAE and RBM. We observed that the RBN delivered consistently better classification results than SAE. The CV and test confusion matrices for RBN with two 120

hidden neurons is shown in Table 3, offering the best performance over other configurations in our tests.

The total accuracy attained by the best RBM for the test step is an impressive 93%, second only to the best 2-layer shallow NN. However, a close look at the confusion matrices for both the CV and the test step reveals the deep net always classifies 1-axle semi-truck (1A) engines to I4 sedan engines in both cases. The high accuracy rate in the test step is because in our collected data there is only one afternoon's data for 1A as test data where the most test data are of I4 sedan engines. The other considerably different behavior of RBM is its low CV accuracy, lower than most classifiers with the exception of NB and SVM (Table 1). This sub-par performance in both CV and test steps of the deep nets should not be viewed as a negative case against this otherwise exceedingly successful learning framework. This is likely caused by the relatively small number of LDV points in the training set by the standard of deep learning. The significantly deeper architecture naturally demands more training data that cannot be met by this pilot study. In this work the shallow NN delivered the best performance and likewise the conventional classifiers such as RF, AdaBoost, and kNN continue to show their enduring ability.

CONCLUSION

In the summer of 2014 the pilot project team supported by the AFRL collected LDV data on 12 different vehicles with four different types of engines. Through intensive exploratory studies we examined and summarized the special phenomenology exhibited by the LDV data in the periodic vibrations of the engine and their multiple modes at the vehicle surfaces and the distinction to human speaker/speech signals. Consequently a new indexing scheme, which differs from conventional speech or music encoding, denoted by the *spectral tone-pitch vibration indexing scheme*, was developed. This indexing scheme takes advantage of LDV's special behavioral properties through performing a Fourier transform, band-pass filtering, data quantization, and finally another Fourier transform to formulate a 120-dimensional vector where the content is 1.25 seconds of overlapping 30 second LDV data. After an intensive model search of more than 10 different typical classifiers and parameter retuning, it was revealed that the 2-layer neural network with 20 hidden neurons delivers the best performance of more than 96% accuracy rates for all four different types of engines investigated.

The outstanding classification results attained by NN is unlikely caused by over-fitting because the training data consists of only approximately 80 LDV points measured on a single day for each engine while the test data is of approximately 2000 points scattered over the entire five week collection. The other classifiers using the same training and test configurations cannot achieve similar performances, as evidenced by Tables 1, 2 and Figure 4. We consider the enabling cause of this network to be the use of 20 neurons in the hidden layer which reduces the dimensionality from 120 to 20. Although it is impossible to precisely decide the correct number of dimensionalities for the manifold it must be considerably smaller than 120. By reason of the *curse of dimensionality*, almost all classifiers, including the celebrated Random Forest and AdaBoost, cannot yield acceptable accuracies. So we conjecture that a dimensionality reduction step needs to be inserted prior

to applying these classifiers to achieve performance similar to the learned NN. This is a research direction to be conducted in the near future.

In our experiments, preliminary results were reported on the use of deep nets to classify the four engines and the performance is shown to lag in the best shallow nets. This is likely caused by the relatively small number of data in the training set that is demanded by the adequate training of a deep net. In order to gain deeper insights more work, i.e., more data collection and algorithmic investigations will be performed in the future to conclusively determine if deep nets can play a role in the effective classification of LDV data to replicate their success to other machine learning tasks (Deng & Yu, 2014).

In this work, we used the majority rule to classify a 30-second data segment L by aggregating the classification labels of all the overlapping smaller segments, the label that receives the most votes is assigned to the entire segment. This majority rule has its own flaws as the winning label may win most contests by slim margins but lose smaller number of contests significantly. If the posterior probabilities of all classifiers are available, the label of the Maximum a Posteriori (MAP) within the Bayesian statistics framework could be a viable choice to determine L 's overall label. This Bayesian approach assumes all posterior probabilities generated by different classifiers are equivalent and thus comparable, which could be a concern, especially if a committee of diversified classifiers is employed. This is future work that needs to be investigated. Another issue with the methodology employed in this paper is the 30-second duration of each LDV measurement in our pilot study, in practical situations it is nearly impossible to collect such long duration data for a moving vehicle. Currently we are actively conducting more data collections to address issues for shorter duration data and moving vehicles based on different distance measure (Wei, 2014) and indexing techniques (Wei, 2013).

Acknowledgments

The authors would like to thank Dr. Vincent Velten as AFRL/RYA's academic advisor in this summer pilot study. The authors would also like to thank fellow team member Mr. David Drewniak of Univ. of Illinois at Urbana-Champaign for his hard work as the designated vehicle driver throughout the project in the summer. We also want to thank Mr. Jim Leonard, Ms Ashley Smith, Dr. Ed Watson and Mr. Steven Goley for valuable academic discussions related to this work. Our thanks also go to Mr. Darrell Barker and Mr. Hyatt Baker (AFRL/Ryaa), Mr. Keven Sigmund (Etegent Inc.) and all AFRL and Etegent researchers who helped us make the data collection process extremely smooth.

References

- Averbuch A, Hochman K. A Diffusion Framework for detection of moving vehicles. *Digital Signal Process.* 2010; 20(1):111–122.
- Bellman, RE. *Dynamic programming*. Princeton Univ. Press; 2003.
- Bengio, Y. *Learning Deep Architectures for AI* (Foundations and Trends in Machine Learning, Book 2). Now Publisher Inc; 2009.
- Bishop, C. *Pattern Recognition and Machine Learning*. Springer; 2007.
- Breiman L. Random Forests. *Machine Learning*. 2001; 45(1):5–32.
- Castellini P, Paone N, Tomasini E. The laser Doppler vibrometer as an instrument for non-intrusive diagnostic of works of art: application to fresco paintings. *Optics & Lasers in Engineering*. 1996 May;25:227–246.

- Deng, L.; Yu, D. Deep Learning: Methods and Applications (Foundations and Trends in Signal Processing Series Book 20). Now Publishers; 2014.
- Duda, R.; Hart, P.; Stork, D. Pattern Classification. 2nd. Wiley; 2001.
- Freund Y, Schapire R. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*. 1997; 55:119–139.
- Gonzalez, RC.; Woods, RE.; Eddins, SL. Digital image processing using MATLAB. 2nd. Gatesmark Publishing; 2009.
- Hartmann, W. Principles of Music Acoustics. Springer; 2013.
- Hinton G, Salakhutdinov R. Reducing the Dimensionality of Data with Neural Networks. *Science*. 2006; 313(5786):504–507. [PubMed: 16873662]
- Huang, X.; Acero, A.; Hon, H. Spoken Language Processing: A Guide to Theory, Algorithm, and System Development. Prentice-Hall; 2001.
- Hyvarinen, A.; Karhunen, J.; Oja, E. Independent component analysis. New York: Wiley; 2001.
- Kim, Y. Sound Propagation: An Impedance Based Approach. Wiley; 2010.
- Kinnunen T, Li H. An Overview of Text-Independent Speaker Recognition: From Features to Supervectors. *Speech Communication*. 2010:12–40.
- Kinsler, L.; Frey, A.; Coppens, A.; Sanders, J. Fundamentals of Acoustics. 4th. Wiley; 2000.
- Kubota K. Development of a remote non-contact measurement system combining laser doppler vibrometer and total station for monitoring of structures. The 3rd Int Conf on Structural Health monitoring of Intelligent Infrastructure. 2007
- Lerch, A. An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics. Berlin, Germany: Wiley; 2012.
- Ma W, Xing D, McKee A, Bajwa R, Flores C, Varaiya P. A Wireless Accelerometer-Based Automatic Vehicle Classification Prototype. *IEEE Trans on Intelligent Transportation Systems*. 2013 May; 15(1):104–111.
- Murphy, K. Machine Learning: A probabilistic perspective. The MIT Press; 2012.
- Oppenheim, A.; Schafer, R. Discrete-Time Signal Processing. 3. Prentice Hall; 2009.
- Palm, R. Master Thesis. Technical University of Denmark; 2012. Prediction as a candidate for learning deep hierarchical models of data.
- Petrovich D, Snorrason M, Stevens M. Identifying Vehicles using Vibrometry Signatures. *Proceedings of the 16th International Conference on Pattern Recognition*, Vol. 2002; 3:253–256.
- Preumont, A. Vibration Control of Active Structures: An Introduction. 3. Springer; 2011.
- Rabiner, L.; Schafer, R. Theory and Applications of Digital Speech Processing. Prentice Hall; 2010.
- Rashed A. Fast Algorithm for Noisy Speaker Recognition Using ANN. *International Journal of Computer Engineering and Technology*. 2014:56–65.
- Sigmund K, Shelley S, Bauer M, Heitkamp F. Analysis of vehicle vibration sources for automatic differentiation between gas and diesel piston engines. SPIE 8391, Automatic Target Recognition XXII. 2012
- Smith A, Mendoza-Schrock O, Kangas S, Derking M, Shaw A. Vehicle Classification using Laser-Vibrometry. SPIE DSS 111 Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR V. 2014
- Solomon, D. A Concise Introduction to Data Compression. Springer; 2007.
- Szeliski, R. Computer Vision: Algorithms and Applications. Springer; 2011.
- Tan, PN.; Steinbach, M.; Kumar, V. Introduction to data mining. Pearson Education Inc; 2006.
- Vapnik, V. Statistical Learning theory. New York: Wiley; 1998.
- Viola P, Jones M. Robust real-time object object detection. *Int J of Computer Vision*. 2004; 57(2): 137–154.
- Wang T, Zhu Z, Taylor C. A multimodal temporal panorama approach for moving vehicle detection, reconstruction and classification. *Computer Vision and Image Understanding*. 2013; 117:1724–1735.
- Wei J. Small moving object detection from video sequences. *Int J of Image and Graphics*. 2013; 14(3)
- Wei J. On Markov Earth Mover's Distance. *Int J on Image and Graphics*. 2014; 14(4)

Willemann, D.; Castellini, P.; Revel, G.; Tomasini, E. Int Conf on Vibration measurements by Laser Techniques. Ancona, Italy: 2004. Structural Damage Assessment in Composite Material using Laser Doppler Vibrometry; p. 21-25.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1.

Sample curves in the spectral bands less than 100 Hz for the 4 different engines: Inline-4 (I4) and V engine with 6 cylinders (V6) for sedans, and diesel engines for 1- (1A) and 2-axle (2A) for semi-trucks.

Engine type	Vehicle Pictures
I4 Sedan
1-Axle semi-truck
2-Axle semi-truck
V6 Sedan

Figure 2.
Sample vehicles for the four types of engines used in the tests.

Figure 3.

Flow chart of the LDV training, validation and test algorithm based on spectral tone-pitch vibration indexing scheme and the steps defined in “majority rule”.

Figure 4.

ROC curves of the optimal 2-layer neural network with 20 hidden-layer neurons. Left: CV step; Right: Test step.

Table 1

Accuracy rates for 10 classifiers in CV and test steps.

Methods	SVM	NB	kNN	RF	AdaBoost
CV	74.9	48.1	96.6	95.5	95.8
Test	39.0	60.1	67.7	73.8	75.1
Methods	LgBoost	RbBoost	RsBoost	LPBoost	NN
CV	96.0	89.7	88.9	93.9	97.4
Test	75.3	72.5	74.3	70.3	96.1

Table 2

Confusion matrices for some of the leading classifiers in CV (first row) and Test steps.

	I4	1A	2A	V6	I4	1A	2A	V6	I4	1A	2A	V6	I4	1A	2A	V6
I4	0.97	0.03	0	0	0.91	0.09	0	0	0.92	0.08	0	0	0.94	0.06	0	0
1A	0.11	0.89	0	0	0.04	0.96	0	0	0.06	0.94	0	0	0.03	0.97	0	0
2A	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0
V6	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
I4	0.62	0.05	0.15	0.18	0.69	0.12	0.11	0.08	0.70	0.10	0.12	0.08	0.97	0.01	0	0.02
1A	0.45	0.55	0	0	0.33	0.67	0	0	0.11	0.89	0	0	0.04	0.96	0	0
2A	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0
V6	0	0	0.02	0.98	0	0	0	1	0	0	0	1	0	0	0	1
	kNN				Random Forest				AdaBoost				2-layer NN			

Table 3
Confusion matrices and accuracies of RBN with two hidden layers and 120 neurons.

	I4	1A	2A	V6	I4	1A	2A	V6
I4	1	0	0	0	I4	1	0	0
1A	1	0	0	0	1A	1	0	0
2A	0	0	1	0	2A	0	0	1
V6	0.22	0	0	0.78	V6	0.19	0	0.04
CV accuracy: 76%				Test accuracy: 93%				