A thesis for the Degree of Doctor of Philosophy

Sound Detection and Classification Algorithms

based on Harmonic Characteristics

Juho Kim

Department of Ocean System Engineering

GRADUATE SCHOOL

JEJU NATIONAL UNIVERSITY

2015. 12.

Sound Detection and Classification Algorithms based on Harmonic Characteristics

Juho Kim

(Supervised by professor Chong Hyun Lee)

A thesis submitted in partial fulfillment of the requirement
for the degree of Doctor of Philosophy

2015. 12.

This thesis has been examined and approved by

Thesis director, Jinho Bae, Professor, Dept. of Ocean System Engineering

----
Chong Hyun Lee, Professor, Dept. of Ocean System Engineering

Seongil Kim, Researcher, Agency for Defense Development

Kyungsup Kim, Professor, Chungnam National University

Youngchol Choi, Researcher, Korea Research Institute of Ships and Ocean
Engineering (KRISO)

2015. 12. 11

Date

Department of Ocean System Engineering
GRADUATE SCHOOL
JEJU NATIONAL UNIVERSITY

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Detection and classification of unmanned aerial vehicles (UAV) have been demanded to reduce many problems such as ground or midair collisions, privacy and security. It is difficult to detect UAV for radar since it is relatively small and flying at low altitude. However, sounds of engine and propeller can be measured easily by microphone. In this thesis, UAV sound detection and classification algorithms are proposed, which are based on harmonic spectral peaks of UAV. When two harmonic spectral peaks are mapped into a vector, it has a unique angle according to its harmonic orders. By pre-defining reference vectors based on known harmonic orders of target UAV and computing similarity of input vectors with the reference vectors, the UAV sound can be identified from non-harmonic sound. Inner product based algorithm and virtual array based algorithm are proposed to compute the similarities. Both algorithms show stable discrimination performance than Mel-frequency cepstral coefficient (MFCC). The under area of ROC curve was up to 0.975 by using the virtual array based algorithm. Harmonic order detection algorithm for unknown signal is also studied. Probabilities of input vectors to be certain reference vectors are computed for all possible cases and extreme learning machine (ELM) are used as a classifier. By concatenating the order detection algorithm and another ELM classifier, multiclass classification can be possible. It shows better performance than MFCC for four different classes of measurement sound data. Proposed algorithms can be usefully applied for

detection and classification of other harmonic sounds as well as UAVs.

Chapter 1

INTRODUCTION

1.1 Background and previous study

An increasing demand for usefulness of unmanned aerial vehicles (UAV) operation has been followed by growing number of the UAV in recent years.[1] Global spending on commercial and civil UAVs reaches 550 million U.S. dollar in 2013. It has doubled in 6 years.[2] When military UAV is included in the survey, the global market can be estimated much larger than this. Unfortunately, there are negatives to the increased usage of UAV in both military and civil area. The UAV can pose significant threat such as ground collision, midair collision, and terrorism. Issues of invasion of privacy and security have been also emerged recently.[3,4] Therefore, prevention measures of these problems has been demanded, such as safety regulations for UAV operation. To control and manage UAV operation to be safety, in more practical point of view, the UAV should be detected and classified from the other objects.

There are broad spectrum in UAV types as shown in Fig. 1.1. The following classes of UAVs are defined and primarily differentiated by mass: Micro, Mini, Tactical, Medium Altitude, and High Altitude/ UCAV (Unmanned Combat Air Vehicle).[3] Among them, the high altitude and massive UAV can be detected by existing radar system. However, even though the market growth is mostly focused on micro and mini UAV, it is still difficult to detect the small-size, low-flying aircraft by using radio frequency (RF), electro-optical (EO),
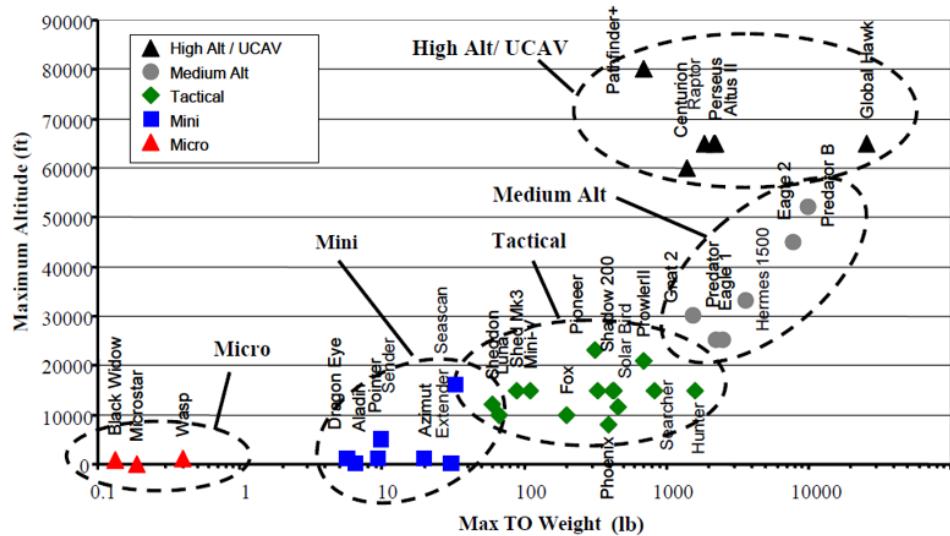
Figure 1.1 spectrum of UAV according to weight and operation altitude [3]
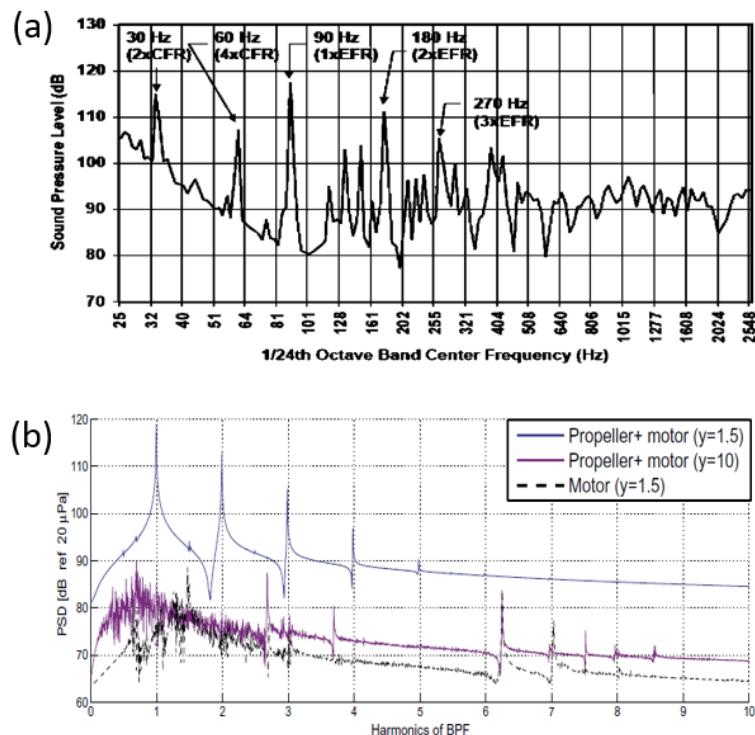
Figure 1.2 (a) Harmonic spectrum generated by engine.[6] (b) Harmonic spectrum generated by propeller.[8]

and infrared (IR) due to weak signal strength and high implementation cost. [5] However, the

sound emitted by engine and propeller of UAV can be easily measured by a microphone. The

engine sound is composed of both cylinder firing rate (CFR) and engine firing rate (EFR) as

shown in Fig. 1.2. CFR is corresponding to the revolutions per minute (RPM) of engine and

ERF is the CFR times the number of cylinder. [6,7] Both CFR and EFR with their harmonics

contributes to form one harmonic spectrum. On the other hand, propeller sound has blading

passing frequency (BPF) which is related on the engine RPM as well as the number of

propeller blades. BPF which is RPM times the number of blades also contributes to generate

harmonic peaks. [8] Therefore, both sound of engine and propeller gives rise to an enhanced

harmonic spectrum according to engine RPM. The harmonic spectrum can be regarded as a

significant sound characteristics of UAV.

There have been research to find direction of arrival (DOA) of the UAV using

microphone array. [9] It was possible to find DOA using conventional method but it is limited to

detect and classify UAV from the other sounds. Only a few researches tried to classify the

sounds by using the characteristics of harmonics. A harmonic line association (HLA) is one of

feature extraction algorithms for harmonic sound classification. [10,11] It is greedy search based

algorithm to find a best matched template subject to range constraint of fundamental

frequency sound as shown in Fig. 1.3. It is very innovative and useful ideal to deal with

harmonic sound. However, it requires expectation of fundamental frequency range, which is

not practical due to not only unforeseeable motions yielding RPM variation but also diverse

types of UAV.
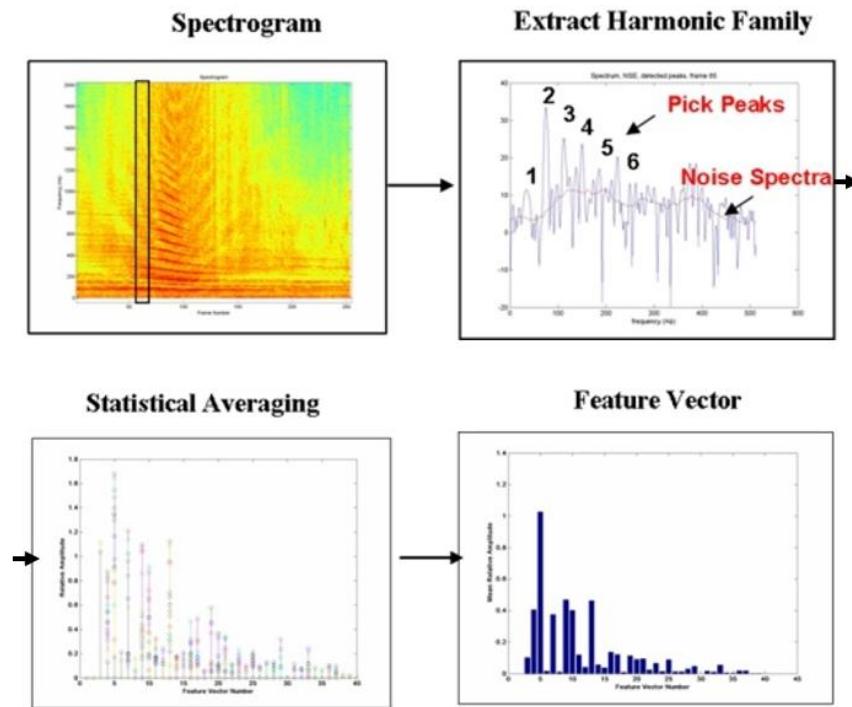
Figure 1.3 Extraction process of HLA

1.2 Objective

To overcome this limit, we suggest new feature extraction algorithm does not require any fundamental frequency information and greedy search. By adopting vector representation, input spectral peaks are converted into two dimensional vectors. Since any sets of two harmonic spectral peaks form unique directional vectors regardless of its fundamental frequency, harmonic sound can be detected and classified by measuring and comparing the vector directions of input signal.

The first goal of this thesis is to identify whether unknown input signal is harmonic or non-harmonic. The achievement of this binary classification could be not only useful information for harmonic sound detection but also foundation of multiclass classification. The second goal is to recognize the harmonic orders of unknown signals. Moreover, types of the harmonic signals are to be classified by the proposed multiclass classification algorithm.

1.3 Thesis layout

Thesis outline is as follows: Basic concept of frequency vectors and its angles are introduced in Chapter 2. A reference vector and a reference angle are also defined. Then the problem statement is followed. Two proposed algorithms are described and those are evaluated using both simulation data and measurement data.

Chapter 3 deals with the harmonic order detector which gives the harmonic order information of unknown signals. By generalizing and integrating the multi-reference angle problem described in previous chapters, it is proved that the harmonic order can be identified using both simulation data and measurement data.

Multiclass classification algorithm is proposed in Chapter 4. By applying harmonic order detector algorithm and using two stage extreme learning machine (ELM) classifier, spectral magnitude of each harmonic order can be estimated and it is used as feature vectors for the multiclass classification. The proposed algorithm is also evaluated using four-class measurement data.

Lastly, this study is concluded with discussion and suggestion for future work.

Chapter 2

BINARY CLASSIFICATION

2.1 Problem statement

In order to identify a harmonic signal from the unknown signals, it is reasonable to analyze a relationship of spectral peaks. Because they are equally spaced on the frequency axis according to the fundamental frequency while that of non-harmonic spectral peaks are not as shown in Fig. 2.1 (a) and (b). In addition, for the realistic signals, small variations on the spectral peaks can be considered including some measurement errors. Then, the harmonic spectral peaks can be modeled as follows:

$$f_i = n_i f_0 + \Delta f, \quad f_j = n_j f_0 + \Delta f \tag{2.1}$$

,where f_0 and Δf are fundamental frequency and variation of spectral peak and n_i and n_j are integer orders of harmonics.

A vector which is composed of f_i and f_j in a two-dimensional plane can be considered as shown in Fig. 2.1 (c). It is defined as

$$\boldsymbol{r}_m = [f_i \quad f_j]^T, \tag{2.2}$$

where $m = 1, 2, \ldots, M$ and M is the number of vectors which are all possible combinations of f_i and f_j. Then the angle of the vector r_m is defined by

$$\theta_m = \tan^{-1}\left(f_j \big/ f_i \right) = \tan^{-1}\left(\frac{n_j f_0 + \Delta f}{n_i f_0 + \Delta f} \right) = \tan^{-1}\left(\frac{n_j + \Delta f / f_0}{n_i + \Delta f / f_0} \right). \tag{2.3}$$

If Δf is assumed to be much smaller than the f_0 so that $\Delta f / f_0$ is approach to zero, θ_m becomes unique constant value with respect to n_i and n_j. It is referred to *reference angle* in this thesis which is defined as

$$\theta_{ref} \equiv \tan^{-1}\left(n_j / n_i\right) \tag{2.4}$$

and a vector composed of the harmonic orders n_i and n_j is called as *reference vector* which is defined as

$$\boldsymbol{r}_{ref} \equiv [n_i \quad n_j]^T. \tag{2.5}$$

Assume that we already know that a target signal has the harmonic orders of n_i and n_j. Then, by comparing the pre-defined $\boldsymbol{r}_{ref}$ or θ_{ref} with $\boldsymbol{r}_m$ or θ_m extracted from input unknown signal, we can identify whether the input signal is harmonic or non-harmonic. If an input signal has harmonic spectrum, $\boldsymbol{r}_m$ or θ_m are well matched with $\boldsymbol{r}_{ref}$ or θ_{ref}
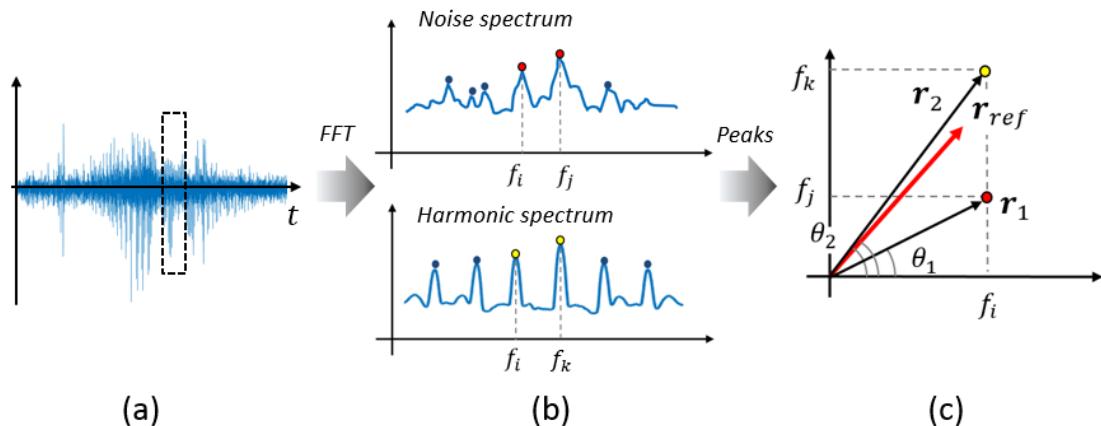
Figure 2.1 (a) A waveform of unknown input signal. (b) Frequency spectrums and their peaks of noise spectrum and harmonic spectrum. (c) Vector representation of spectral peaks

and their distance is closer than the others which are from non-harmonic spectrums. However, if an input signal has non-harmonic spectrum such as ambient noise, the distance between them is larger than that of harmonic signals.

Therefore, the problem is to classify target signals and noise using statistical characteristics of a distance between r_{ref} or θ_{ref} with r_m or θ_m. Two distance measurement algorithms are proposed in this thesis. The first is to compare r_m with r_{ref} and the second is to compare θ_m with θ_{ref}.

2.2. Inner product based algorithm

2.2.1. Algorithm

Let $F(n)$ be a set of peak frequencies at n^{th} frame of input signal as shown in Fig. 2.2 (a). It can be written as follow:

$$F(n) = \left\{\ f_q(n)\ \middle|\ q = 1, \dots, i, \dots, j, \dots, Q\ \right\}, \tag{2.6}$$

where Q is the number of dominant peak frequencies. When $F(n)$ is obtained from harmonic signal, then two components of $F(n)$ can be modeled as

$$f_i(n) = n_i f_0(n) + \Delta f(n), \quad f_j(n) = n_j f_0(n) + \Delta f(n). \tag{2.7}$$

Assume that Δf can be modeled as zero mean normal distribution with standard deviation σ_f. Meanwhile, two component of $F(n)$ obtained from non-harmonic ambient noise can be modeled as

$$f_i(n) \sim U\left(0, \frac{1}{2}f_s\right), \quad f_j(n) \sim U\left(0, \frac{1}{2}f_s\right) \tag{2.8}$$

where f_s is sampling frequency and $U(a, b)$ represents uniform distribution of interval a and b.

The key idea of this algorithm is to compute similarity between input frequency vectors and the reference vector. If we use the vectors $\boldsymbol{r}_m$ and $\boldsymbol{r}_{ref}$ introduced by Eq. (2.2) and (2.5), the influence of f_0 cannot be avoided. Therefore, both input frequency vectors and the reference vector should be normalized newly defined as

$$r_m = \eta[f_i \quad f_j]^T, \quad m = 1, 2, \dots, M, \tag{2.9}$$

10
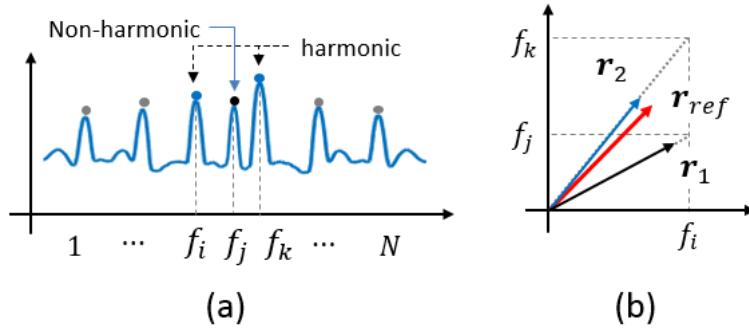
Figure 2.2. (a) Examples of frequency spectrum and its peaks. (b) Vector representation for combinations of spectral peaks. Note that the vectors are normalized to be compared.

and

$$r_{ref} = \eta_{ref}[n_1 \quad n_2]^T \ , \tag{2.10}$$

where M = $_NC_2$ (2-combination from N peak frequencies). η and η_{ref} are the normalization factors of r_m and r_m, respectively, and they are defined as $\eta = \left(f_i^2 + f_j^2\right)^{-1/2}$, and $\eta_{ref} = (n_1^2 + n_2^2)^{-1/2}$. As mentioned before, pre-defined r_{ref} is used as a reference to find the closest vector among M of r_m. Fig. 2.2 (b) shows that the vector r_2 from harmonic spectral peaks has higher similarity with r_{ref} than r_1 from non-harmonic peaks.

2.2.1.1. Decision based on single frame

Let ω_0 and ω_1 are the class of harmonic signal and white Gaussian noise. Then, the similarity between the reference vector and input vector can be measured by square of scalar

product of them, which is denoted by

$$\begin{cases} \omega_0 \colon w_M(n) = \max\langle r_{ref}{}^T r^0{}_m(n)\rangle^2 \\ \omega_1 \colon w_M(n) = \max\langle r_{ref}{}^T r^1{}_m(n)\rangle^2 \end{cases} \tag{2.11}$$

where $r^0{}_m(n)$ and $r^1{}_m(n)$ are normalized spectral peaks vectors of non-harmonic noise and harmonic signal, respectively. Then we suggest to decide ω_1 by Bayesian classifier using log likelihood y defined as

$$y(x) = \log(p(x|\omega_1)p(\omega_1)) - \lambda \log(p(x|\omega_0)p(\omega_0)) > 0 \tag{2.12}$$

where $x = w_M$.[15,16] Suppose that both priori probability $p(\omega_0)$ and $p(\omega_1)$ are the same and λ is 1 for the simplification. Both likelihood $p(x|\omega_0)$ and $p(x|\omega_1)$ follows gamma distribution with shape parameter α_z and rate parameter β_z, which are described in appendix A. Then it can be written as

$$p(x|\omega_z) \sim \text{Gamma}(\alpha_z, \beta_z), \quad z = 0 \text{ or } 1 , \tag{2.13}$$

where $\alpha_0 = \hat{\alpha}_n$, $\beta_0 = \hat{\beta}_n$ and $\alpha_1 = 1/2$, $\beta_1 = 2\sigma_f/(n_1{}^2 + n_2{}^2)f_0{}^2$. $\hat{\alpha}_n$ and $\hat{\beta}_n$ are gamma parameter estimation values by simulation of noise data generation. Their PDF is shown in Fig. 2.3 (a).

2.2.1.2. Decision based on multiple frames

Assuming that input harmonic sound can be measured for N frames, mean and standard deviation of $\{w_M(1), w_M(2), \dots, w_M(N)\}$ can be used for better performance on

classification. The distribution of mean w_μ and standard deviation w_σ corresponding to both

target harmonic sound and ambient noise can be written as

$$w_\mu = \frac{1}{N} \sum_{n-N}^{n} w_M(n) \sim \text{gamma}(N\alpha_z, N\beta_z)$$

$$w_\sigma = \frac{1}{N-1} \sum_{n-N}^{n} \left(w_M(n) - w_\mu\right)^2$$

$$\sim \text{Generalized gamma}\left(2, \frac{\alpha_z}{\beta_z{}^2}, \frac{\alpha_z}{\beta_z{}^4}\left\{\frac{2N\alpha_z}{(N-1)^2} + \frac{6}{N}\right\}\right), \qquad z = 0 \; or \; 1 \qquad (2.14)$$

Derivation of Eq. (2.14) are presented in Appendix A.

Shape parameter and rate parameter of new gamma distribution is changed according to

number of samples N in both w_μ and w_σ so that the advantage on discrimination of target

signals from ambient noise can be given as shown in Fig. 2.3 (b) ~ (d). Now we can rewrite

$x = \left[w_\mu \; w_\sigma\right]^T$ of Eq. (2.12).

Until now, only a single reference vector $\boldsymbol{r}_{ref}$ has been considered but it is easily

extended to the multiple reference problem to include additional reference vectors. A block

diagram for the proposed algorithm is shown in Fig. 2.4.

13
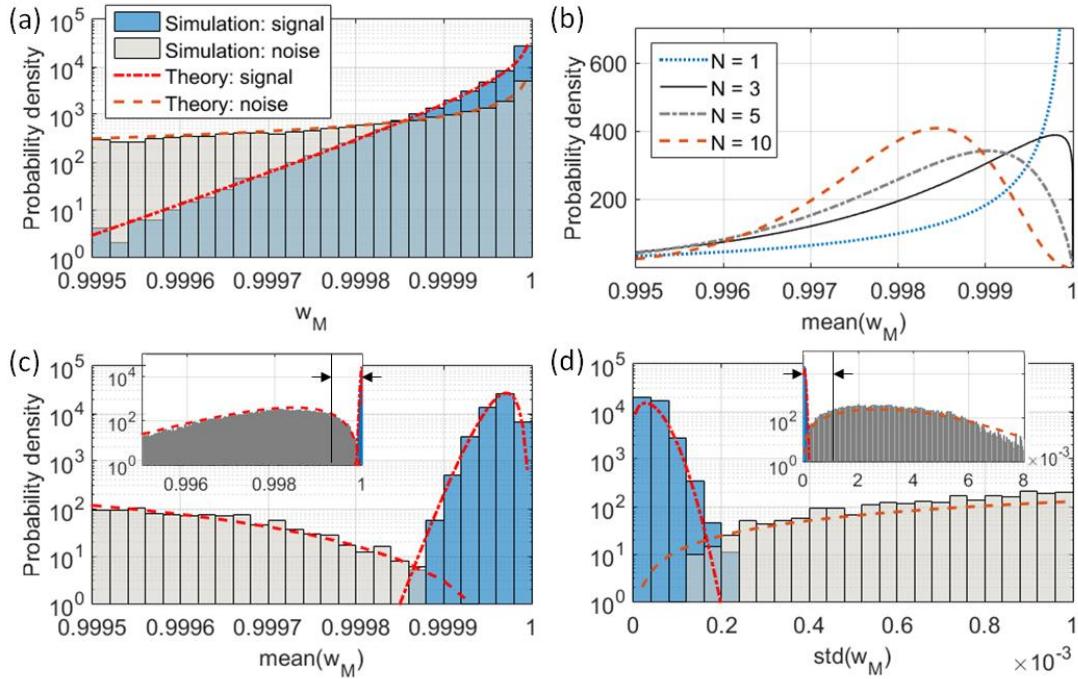
Figure 2.3 (a) distribution of w_M of harmonic signal and white Gaussian noise. (b) Variation of PDF with number of frames N. (c) PDF of w_μ for $N = 10$. (d) PDF of w_σ for $N = 10$
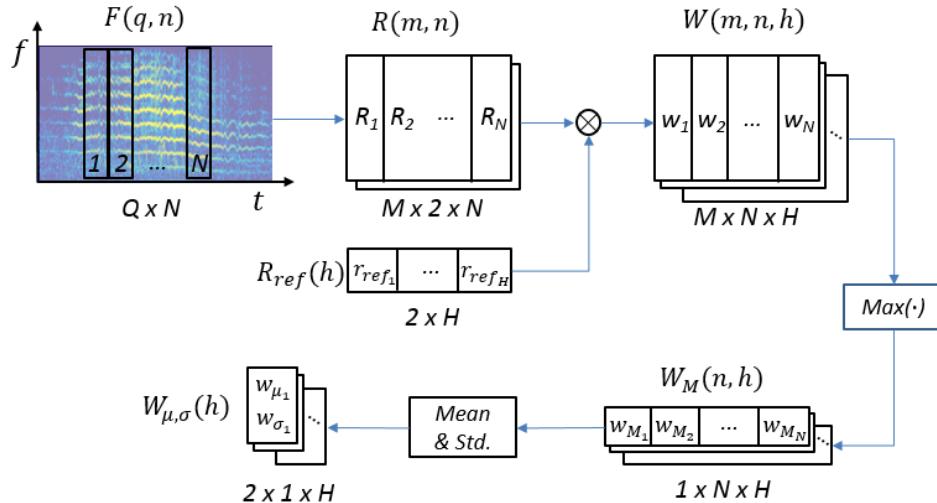
Figure 2.4. Block diagram for multiple frames and multiple references

2.2.1.3. Joint distribution estimation and classifier

In order to consider both w_μ and w_σ for x, joint probability should be considered. However, theoretical bivariate generalized gamma distribution is not appropriate for the real time detection algorithm due to its iterative computation.[22] Therefore, Gaussian mixture model (GMM) is used for estimation of joint probability function,[14] which is written as

$$p(x|\omega_z) = \sum_{t=1}^{T} c_t \, p(x|\mu_t, \sigma_t) \,. \tag{2.15}$$

Fig. 2.5 (a) and (b) show joint pdf of w_μ and w_σ by simulation and GMM estimation, respectively. Considering that $x_h = \left[w^h{}_\mu \; w^h{}_\sigma\right]^T$, which is random variable with respect to h^{th} reference, we can easily carry out multi-order detection. Since $x_1, \ldots, x_H$ are independent where H is number of reference, Naive Bayesian classifier can be applied as bellow [14]

$$p(x|\omega_z) = \prod_{h}^{H} p(x_h|\omega_z) \,, \mathrm{z} = 0 \; or \; 1 \tag{2.16}$$

Then, by using likelihood in Eq. (2.16), detection error is computed as follows:[15]

$$P_e = p(\omega_1|\omega_0)p(\omega_0) + p(\omega_0|\omega_1)p(\omega_1). \tag{2.17}$$
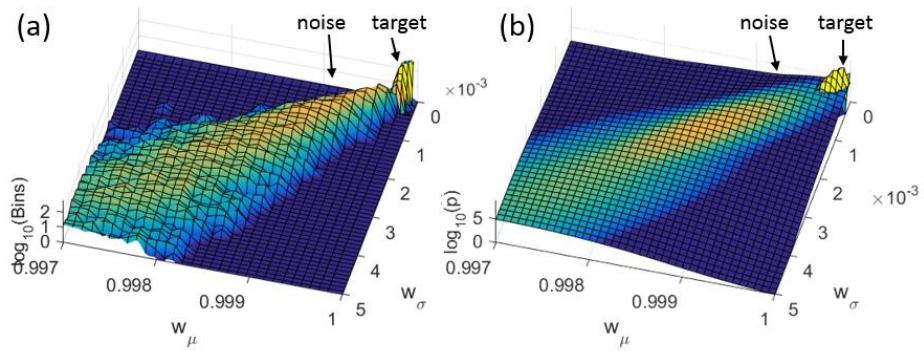
Figure 2.5. (a) Probability density by simulation using $\sigma_f = 3, Q = 7$ and $r_{ref} =$ $[300\ 400]^T$. (b) Probability density function by Gaussian mixture model. Number of Gaussian mixture T is 3.

2.2.2 Results

2.2.2.1. Simulation data

Theoretical errors by Eq. (2.12) is computed using simulation data and they are shown in Fig. 2.6 ~ Fig. 2.8. Spectral peaks were modeled by Eq. (2.7) and Eq. (2.8). f_i and f_j are randomly generated in 10,000 times and they are used for input values of the proposed algorithm.

Error according to Q is shown in Fig. 2.6 when $n_2/n_1 = 1.333$ and $\sigma_f = 3$. The error rate of single frame and single reference shows 18% ~ 38% but it sharply increase while the number of frames and references increase. When ten spectral peaks are selected for this algorithm where $N = 10$ and $H = 3$, 2.5% of error rate can be expected.

n_j/n_i also affects to the error rate as shown in Fig. 2.7. It is computed when $Q = 10$ and $\sigma_f = 3$. This results shows that the error rate can be controlled by selection of n_2/n_1.

Fig. 2.8 shows the error rate according to σ_f when $n_2/n_1 = 1.333$ and $Q = 10$. Smaller error rates can be expected when σ_f decreases.
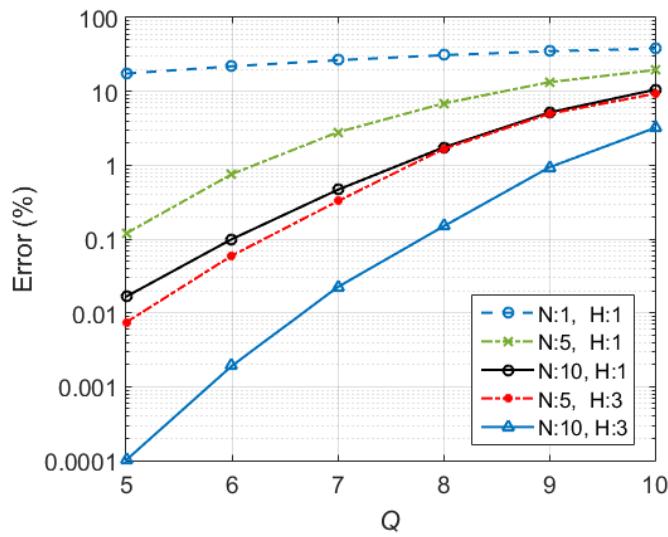
Figure 2.6 Error rate according to number of peaks Q
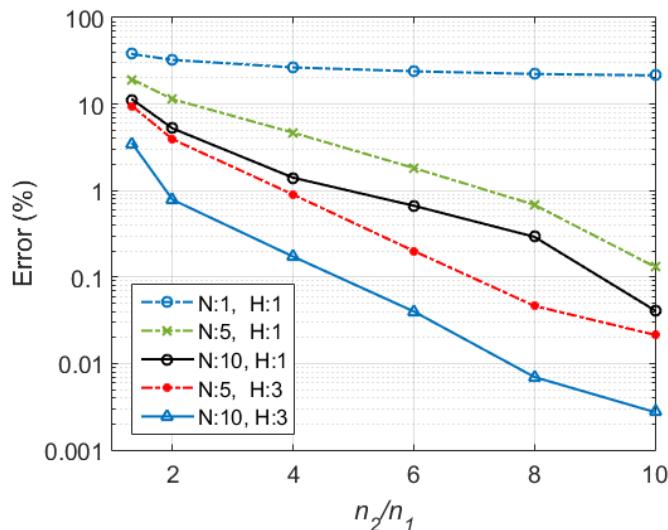
Figure 2.7 Error rate according to n_j/n_i
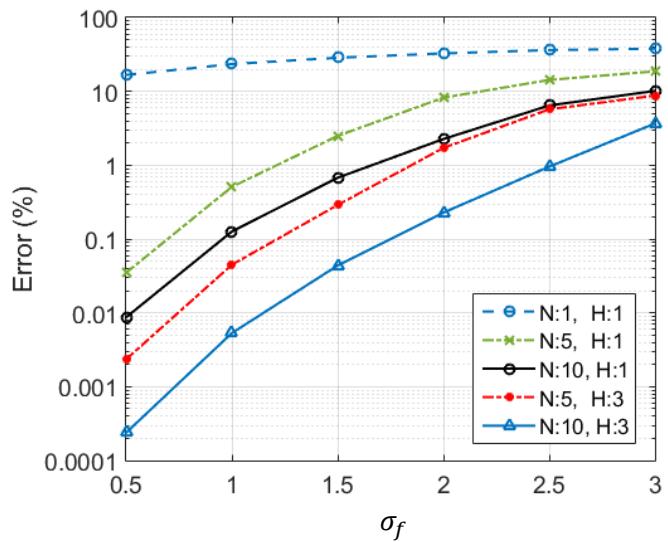
Figure 2.8 Error rate according to σ_f of target signal

2.2.2.2. Measurement data

To show feasibility under real environment, the sound of flying UAV was recorded, which is equipped four cycle gasoline engine and flies back and forth near a microphone at 150 altitude. Sampling frequency was 1024 Hz and noise whitening with band pass filtering of 50~500 Hz was applied for preprocessing. For the comparison, ten Mel-frequency cepstral coefficients (MFCC) with twenty Mel-filter banks are extracted from the training data and GMM is applied for estimation of bivariate Gaussian pdf. Twenty data of UAV sound and ambient noise whose duration is 10 sec are used for training with 0.5 sec of window and 50 % of overlap.

Fig. 2.5 (a) shows spectrogram of test data. UAV was approaching from 40 sec and it passed by the microphone t_2 and it moved away at t_3. Before the UAV reached at measurement site, a car passed by the microphone at t_1. Frequency spectrum corresponding to each time point $t_1, t_2,$ and t_3 is depicted in Fig. 2.5 (b).

y computed by MFCC becomes incorrectly positive during 8 sec at t_1 because of the car sound as shown in Fig. 2.5 (c). On the other hand y gives negative values even though UAV sound continued at t_2 and t_3, when the sound energy reduces. However, y computed by proposed method stably shows negative values even when car passed by the microphone at t_1.

Since 8th order of harmonic peak disappears from 80 sec, y computed by $n_1 = 5$ and

$n_2 = 8$ becomes negative at 80 sec. However, it still remain positive when y computed by

$n_1 = 5$ and $n_2 = 7$ due to existence of harmonic orders. y computed by multi-reference

shows more stable results than that of single reference.
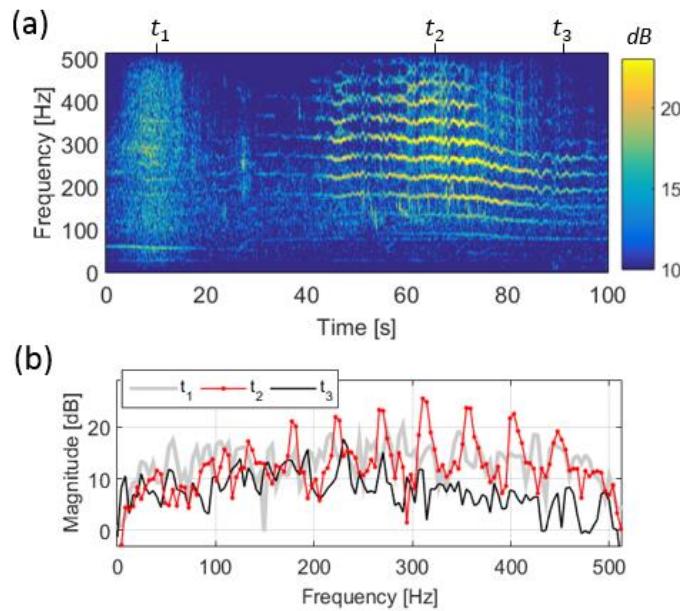
Figure 2.9. (a) Spectrogram of measurement sound of UAV, (b) frequency spectrum at

t_1, t_2, t_3 marked in Fig. 6 (a)
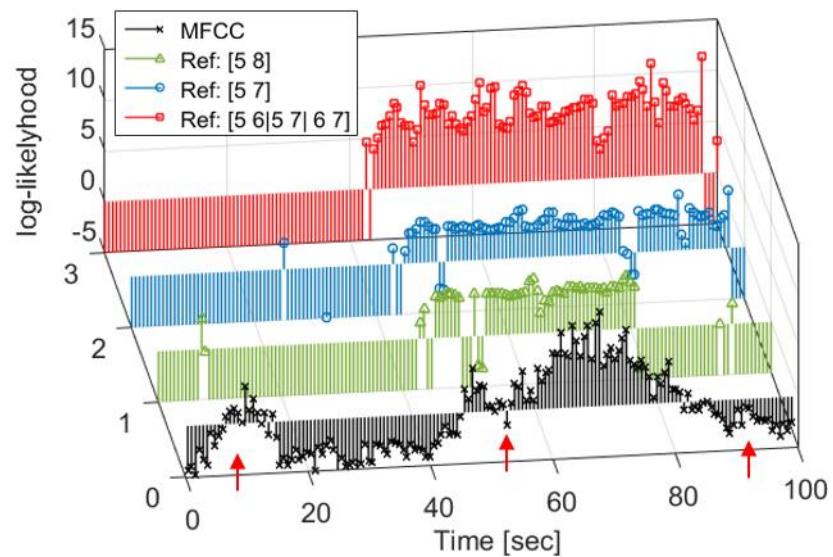
Figure 2.10 Detection results from proposed method and MFCC

2.3. Virtual array based algorithm

2.3.1. Algorithm

The second algorithm is to measure a similarity between θ_m and θ_{ref}. Basic idea is

the same with last chapter as shown in Fig. 2.11 (a) and (b). As mentioned previously, the

advantage of this idea is to remove the influence of f_0 to find harmonics. In addition to this,

vector normalization which is used in last chapter does not need to compare θ_m with θ_{ref}

as shown in Fig. 2.11 (b). In order to measure a similarity between two angles, we adopted

array signal processing. In this point of view, θ_{ref} and θ_m are regarded as a desired

direction and input signal direction. Let us $X(\theta_{ref},\theta_m)$ be a similarity function between

θ_{ref} and θ_m. Then, by following conventional array beamformer equation, we define

function $X(\theta_{ref},\theta_m)$ as follows:
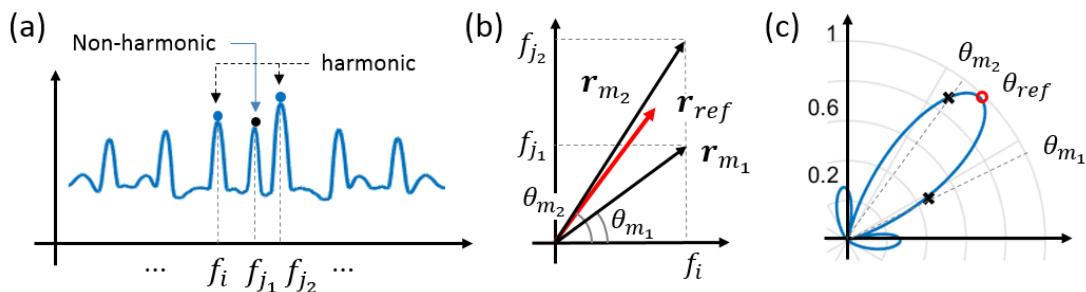
Figure 2.11 (a) Example of harmonic spectral peaks and non-harmonic spectral peaks (b)
vector representation of spectral peak combination and its angles (c) Example of similarity
function $X(\theta_{ref},\theta_m)$ by beam pattern

$$X(\theta_{ref}, \theta_m) = w(\theta_{ref})^H a(\theta_m) \qquad (2.18)$$

, where $a(\theta_m)$ is array response vector and $w(\theta_{ref}) = a(\theta_{ref})$.

For simplicity, uniform line array (ULA) with L elements is considered so that $a(\theta_m) = [1, e^{j\pi \sin \theta_m}, \dots, e^{j(L-1)\pi \sin \theta_m}]^H$ where H represents complex conjugate transpose[12]. Note that $X(\theta_{ref}, \theta_m)$ yields maximum output when $\theta_m = \theta_{ref}$ as shown in Fig 2.11 (c). At n^{th} frame, beamformer output $y(n)$ can be written as

$$y(n) = w(\theta_{ref})^H u(n) \qquad (2.19)$$

, where $u(n) = \sum_{m=1}^{M} a(\theta_m(n))$ [12]. To handle with K reference angles of $\theta_{ref_1}, \dots, \theta_{ref_K}$, we use LCMV beamformer described as follows

$$\min_{w} w^H R_u w \quad \text{subject to } Cw = g \qquad (2.20)$$

, where $R_u = E\{u(n)u^H(n)\}$, $C = [a(\theta_{ref_1}) \dots a(\theta_{ref_K})]^H$, $g = [1 \dots 1]^H$ [13]. The well-known optimum weight can be found as

$$w = R_u^{-1} C (C^H R_u^{-1} C)^{-1} g \qquad (2.21)$$

Note that the range of θ_m are $[\pi/4, \pi/2]$ due to the condition of $j > i$. It can be crucial limitation on array processing. Firstly, the PDF of noise is not uniform through the angles. Secondly, θ_m are distributed only limited range so that it is not proper for the LCMV beamforming.

In case of the white Gaussian noise in eq. 2.8, the distribution of θ_m follows $\operatorname{cosec}^2 \theta_m$ which can be obtained by coordinate transformation of random variable f_i and f_j from Cartesian to polar coordinate. It is proved in Appendix A. In order for angle whitening for the

noise, probability integral transform is applied.[17] In addition, the range of θ_m is extended

over $[-\frac{\pi}{2}, \frac{\pi}{2}]$ by following process:

$$\tilde{\theta}_m = \pi(-\cot\theta_m + 1/2). \tag{2.22}$$

where $\tilde{\theta}_m$ is whitened and expanded input angles.

Note that since M is much greater than K at each frame, $u(n)$ can be treated as array

response of multiple scatters and $u(n)$ can be assumed to be independent. Thus, we can

assume $R_u = E\{u(n)u^H(n)\} = I$ and can have new weight vector described as

$$w = C(C^H C)^{-1} g \tag{2.23}$$

where, $C = \left[a(\tilde{\theta}_{ref_1}) \dots a(\tilde{\theta}_{ref_K})\right]^H$ and $g = [1 \dots 1]^H$.

Note that w of Eq. (2.23) can be computed in advance with low complexity. In this paper

we use the beamformer output $y(n)$ of Eq. (2.19) as features with respect to w in Eq. (2.23).

To evaluate detection performance, Bayes decision rule are used, which is described as

follows:

$$C_1 p(y(n)|\omega_1)p(\omega_1) > C_2 p(y(n)|\omega_2)p(\omega_2), \tag{2.24}$$

where C_1 is cost of deciding ambient noise (ω_2) when target signal (ω_1) is true and C_2 is cost

of deciding ω_1 when ω_2 is true.[16] The $p(y|\omega_1)$ and $p(\omega_1)$ are likelihood and prior

probability of UAV signal and $p(y|\omega_2)$ and $p(\omega_2)$ are those probabilities of ambient noise.

2.3.2 Results

2.3.2.1 Simulation data

To show invariance to fundamental frequency, we consider simulation with which fundamental frequency changes from 50 to 120 Hz with fundamental frequency disappears at 10 second as shown in Fig. 2.12 (a). Δf is modeled by zero mean normal distribution with standard deviation σ_f. By using $\tilde{\theta}_{ref_1} = \tan^{-1}\left(2/1\right)$, $\tilde{\theta}_{ref_2} = \tan^{-1}\left(3/2\right)$, $\tilde{\theta}_{ref_3} = \tan^{-1}\left(4/3\right)$, $N = 4$ and $L = 30$, the proposed $y(n)$ is shown in Fig. 2.12 (b). We can observed that the proposed algorithm performs well and is robust to frequency variation.

2.3.2.2. Measurement data

The sound of flying UAV at altitude of 150 m was recorded for 10 seconds with sampling frequency of 1024 Hz. For preprocessing, noise whitening and band pass filtering of 50 ~ 500 Hz was applied. For training data, we use ten UAV data set and ten ambient noise set. For short time Fourier transform, frame length of 512 samples with 50 % overlap is used and seven peak frequencies ($N = 7$) are selected.

For comparison with filter bank based feature, ten MFCCs are extracted by twenty Mel-filter banks. For decision criterion of MFCC, we use difference of Euclidean distance between input MFCC vector $v(n)$ with mean values of signal μ_s and noise μ_n. It is defined
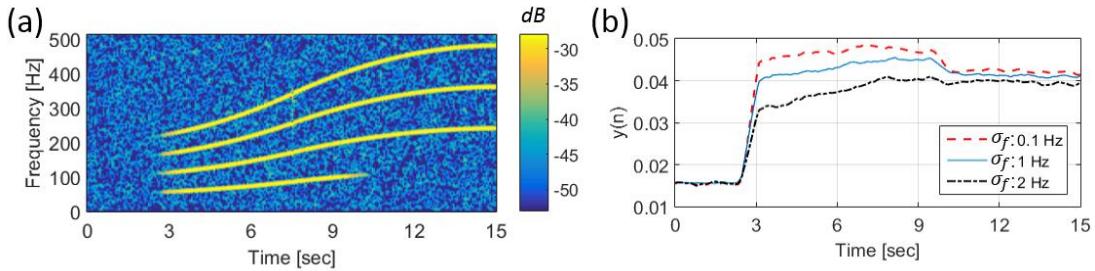
Figure 2.12 (a) Simulated harmonic spectrum whose fundamental frequency changes from 50 to 120 Hz. The fundamental frequency disappears at 10 second. (b) Time series of virtual array beam former output $y(n)$ for Fig. 2 (a), where $Q = 4$, $L = 30$ and $K = 3$ of $\tilde{\theta}_{ref_1} = \tan^{-1}(2/1)$, $\tilde{\theta}_{ref_2} = \tan^{-1}(3/2)$, $\tilde{\theta}_{ref_3} = \tan^{-1}(4/3)$.
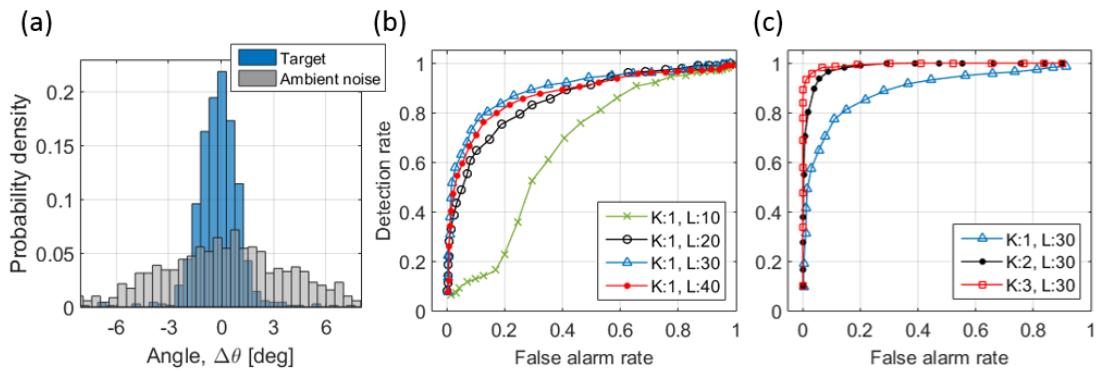
Figure 2.13 (a) Histogram of θ_m's by measurement data set which are closest to θ_{ref}; $[1\ 2]^T$ (b) ROC curve according to number of element L computed using measurement data set (c) ROC curve according to number of element K

as $z(n) = z_n(n) - z_s(n)$, where $z_s(n) = |v(n) - \mu_s|$ and $z_n(n) = |v(n) - \mu_n|$. To obtain

optimum beam width, we investigate distributions of $\tilde{\theta}_m$ from UAV and ambient noise by

using $\tilde{\theta}_{ref} = \tan^{-1}\left(2/1\right)$. The distributions are shown in Fig. 2.13 (a), in which probability

density of θ_m's from UAV are greater than that of ambient noise from -2.25° to 1.75°. With

this information we obtain optimum beam width or the number of array element L=30 by

using half power beam width (HPBW) $\theta_H = 2\sin^{-1}(0.891/L)$ [12].

The influence of L and K on the discriminative performance by $y(n)$ in Eq. (2.19) is

studied by evaluating receiver operating characteristic (ROC) curve as shown in Figs. 2.13 (b)

and (c). Detection performance with $K = 1$, increases as aperture size increases up to $L = 30$

elements and then decreases. The under area a ROC curve (UAC) with $L = 10$, 20, 30,

and 40 are obtained as 0.644, 0.838, 0.878 and 0.862, respectively.[18] The increase of K

yields substantial UAC performance increase up to 0.964 and 0.975.

To show feasibility under real environment, we select test data, which contains sound of car

passing at t_1 and sound of UAV moving closer and passes away for approximately 60 seconds.

The spectrogram of the data is shown in Fig. 2.14(a). The sound of UAV starts at 40 second

and becomes gradually louder until 70 second, t_2 with power and frequency fluctuations.

After 70 second, the sound decreases slowly so that a few low order harmonics are left for last

20 seconds, t_3. The frequency spectrums at t_1, t_2 and t_3 are shown in Fig. 2.14(b). At t_1,

no HSP can be observed and spectrum energy is little smaller than that of t_2 and bigger than

that of t_3. At t_2, HSPs are clearly observed while few peaks are visible at t_3.
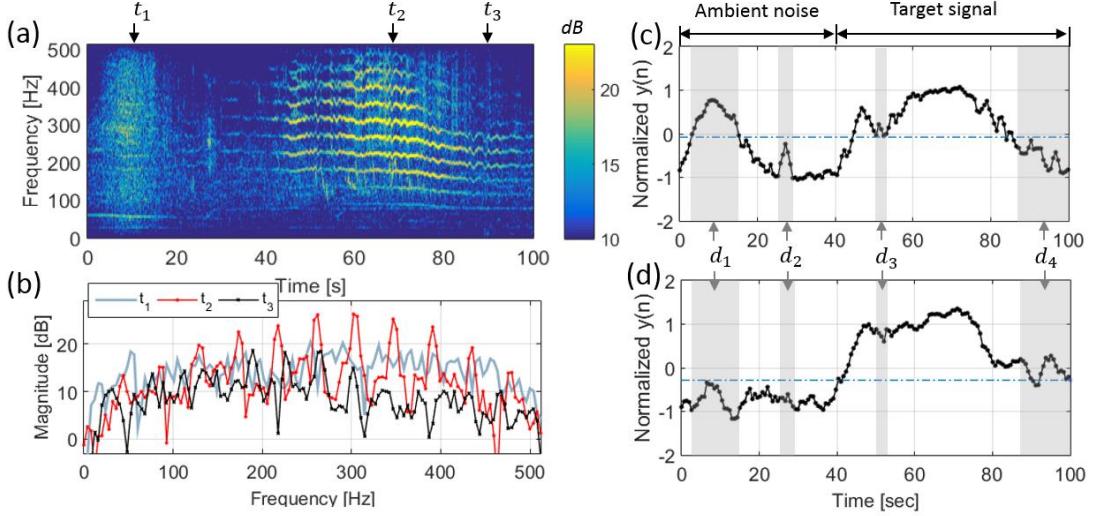
Figure 2.14 (a) Spectrogram of test data. UAV sound appears from 40 second and it emits largest energy at 70 second (t_2). It gradually decreases from 80 second while it loses high order harmonics (t_3). A car sound is ahead of UAV sound from 0 to 20 second (t_1) (b) frequency spectrums at (t_1), (t_2) and (t_3) (c) Normalized difference of Euclidean distance $y(n)$ between MFCC centroid of test data and test data (d) Normalized beam former output $y(n)$ where $Q = 7$, $L = 30$ and $K = 3$ which are $[1\ 2]^T$, $[2\ 3]^T$ and $[3\ 4]^T$.

Normalized $y(n)$s computed by MFCC and the proposed algorithm are shown in Fig. 2.14 (c) and Fig. 2.14(d), respectively. Detection threshold for low incorrect detection is estimated by setting $C_1 = 10$ and $C_2 = 1$ in Eq. (2.24) and is shown as dashed dot line. Shaded area represents noticeable duration marked with duration number $d_1, \dots, d_4$.

Comparing durations of strong ambient noise labeled d_1 and d_2, we can observe that proposed feature shows stable detection while MFCC false detection. Similar performance can

be observed at duration d_3 in which spectrum energy drops. At duration d_4 of low UAV

sound, we can see the proposed algorithm exhibits stable detection while MFCC fails.

Chapter 3

HARMONIC ORDER RECOENITION

3.1 Preprocessing

3.1.1. Cepstral analysis

In the colored noise environment, unwanted noise peaks tend to appear in the particular frequency range. Since it is against the white Gaussian noise assumption, a bad influence can be given in angle whitening and make the performance deteriorate. Hence, we extract whitened spectral peaks from the colored input spectrum by adopting cepstral analysis.[19] It is widely used in speech signal processing to find pitches or formants of human voice.

The cepstrum $c[n]$ is defined as the inverse DFT of the log magnitude of the DFT of signal as follow

$$\begin{cases} X_p[k] = \displaystyle\sum_{n=0}^{N-1} x[n]e^{-j\frac{2\pi}{N}kn} & 0 \le k \le N-1 \\[2mm] \hat{X}_p[k] = \log|X_p[k]| & 0 \le k \le N-1 \\[2mm] c[n] = \dfrac{1}{N}\displaystyle\sum_{k=0}^{N-1} \hat{X}_p[k]e^{j\frac{2\pi}{N}kn} & 0 \le n \le N-1 \end{cases} \qquad (3.1)$$

where $X_p[k]$, $\hat{X}_p[k]$, and $c[n]$ represent DFT, log magnitude of the DFT, and cepstrum, respectively. Those are depicted in Fig. 3.1 (a), (b), and (c). Variable n of $c[n]$ is called as quefrency. Then a cepstral window is applied into the $c[n]$ to lifter out the high quefrency

part of the cepstrum. By inversely transforming the liftered cepstrum, we can get the spectral

peaks (high quefrency part) and spectral baseline (low quefrency part) as shown Fig. 3.1 (d)

and (e).
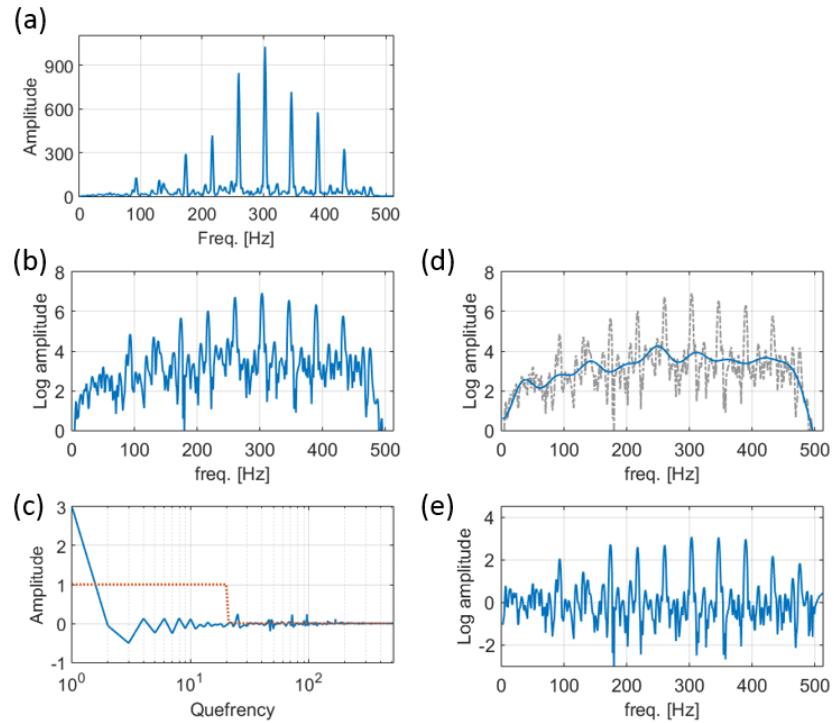
Figure 3.1 (a) Example of input frequency spectrum $X_p[k]$ (b) Log magnitude of DFT $\hat{X}_p[k]$ (c) Cepstrum $c[n]$ (solid) and cepstral window for lifgering high quefrency component (dotted) (d) Inverse transform output of low quefrency components (solid) and log magnitude of DFT of original signal (e) Inverse transform output of high quefrency components

Cepstral decomposition is applied into the UAV signal as examples. Fig. 3.2 (b) and (c) shows that the spectral baseline and spectral peaks, respectively, extracted from measured UAV sound which is shown in Fig. 3.1 (a).
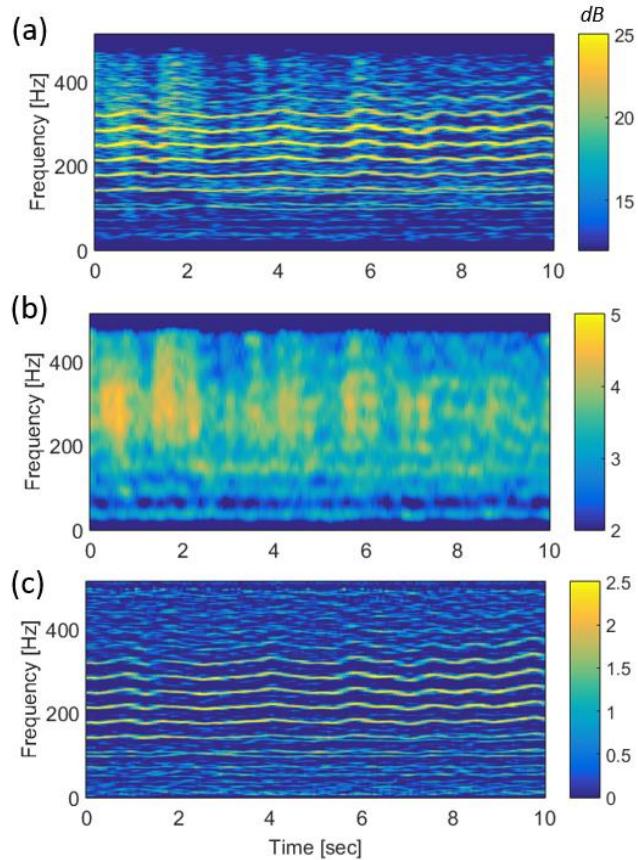
Figure 3.2 (a) Measured UAV sound (b) spectral baseline obtained from low-quefrency cepstral lifting (c) spectral peaks obtained from high-quefrency cepstral lifting

3.1.2. Harmonic to noise ratio (HNR)

HNR is defined as the mean difference between the harmonic spectral peaks $\hat{X}_p(q)$ and the spectral baseline $\hat{X}_b(q)$ at these peak frequencies,[20] which is written as

$$HNR\,[dB] = \frac{1}{Q} \sum_{q}^{Q} 10 \log_{10}(F(q)) \; , \tag{3.2}$$

where $F(q) = e^{(\hat{X}_p(q) - \hat{X}_b(q))}$ and F and Q are DFT amplitude and number of peaks, respectively. In Fig. 3.3, dashed line is spectral baseline and circles and squares are $\hat{X}_p(q)$ and $\hat{X}_b(q)$, respectively. Since $F(q)$ from harmonic spectral peaks has higher values than that from noise peak generally, it can be considered as a weight factor with respect to the frequency vector representation.

Let us introduce the weight γ_m as follows:

$$\gamma_m \equiv F(i)\,F(j) \tag{3.3}$$

where $1 \le i < j \le Q$. There is no weighting for input angles which is from the spectral peaks as shown in Fig. 3.4 (a). By using Eq. (3.3), however, we can apply the weights for the input angles as shown in Fig. 3.4 (b).
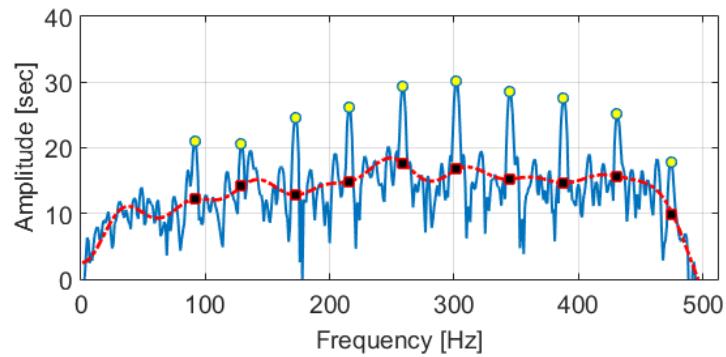
Figure 3.3. Input spectrum of UAV sound (solind) with spectral baseline (dashed). Circles and squares represnet spectral peaks $\hat{X}_p(q)$ and baseline at these peak freqneies $\hat{X}_b(q)$.
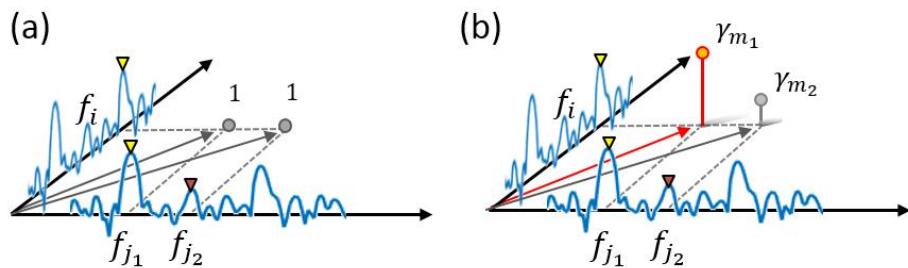
Figure 3.4. Diagrams for (a) no weights and (b) weights based on HNR

3.2 Algorithm

3.2.1 Feature extraction

Suppose that all expanded input angle $\tilde{\theta}_m$ are from the harmonic spectral peaks where $m = 1, \cdots, M$, then each of them can be modeled as follows:

$$\tilde{\theta}_m = \tilde{\theta}_{ref_m} + \Delta\theta_m, \tag{3.4}$$

where $\Delta\theta_m \sim N(0, \sigma_m)$. Because some of $\tilde{\theta}_{ref_m}$ are overlapped when $\tilde{\theta}_{ref_m} = \tan^{-1}(an_j/an_i)$ where $a = 1,2,\cdots$, unique reference angle $\tilde{\theta}_{ref_h}$ should be considered where $h = 1,\cdots,H$ and $H < M$. For example, $\begin{bmatrix} n_i & n_j \end{bmatrix} = \begin{bmatrix} 1 & 2 \end{bmatrix}$ and $\begin{bmatrix} n_i & n_j \end{bmatrix} = \begin{bmatrix} 2 & 4 \end{bmatrix}$ have the same $\tilde{\theta}_{ref_m}$. Then we can rewrite Eq. (3.4) as

$$\tilde{\theta}_m = \tilde{\theta}_{ref_h} + \Delta\theta_m. \tag{3.5}$$

Let ω_h represent h^{th} class corresponding to $\tilde{\theta}_{ref_h}$. Then, the problem is to classify the $\tilde{\theta}_m$ into the correct ω_h.

Proposed algorithm is depicted in Fig. 3.5. Given that we can model the all PDF $f(\tilde{\theta}_m)$ in advance, then input $\tilde{\theta}_m$ can be classified as class ω_h by selecting the maximum probability $P(\omega_h|\tilde{\theta}_m)$ based on the $f(\tilde{\theta}_m)$. Note that $P(\omega_h|\tilde{\theta}_m)$ becomes large value when there are HSPs in the input spectrum according to modeled $f(\tilde{\theta}_m)$ but unwanted noise spectral peak can randomly contribute to the rising $P(\omega_h|\tilde{\theta}_m)$ such as in low SNR environment. Hence, we observe multiple frames and compute averaged probability to decrease the influence of noise peaks. In addition, we adopt the weight γ_m of Eq. (3.3) on to
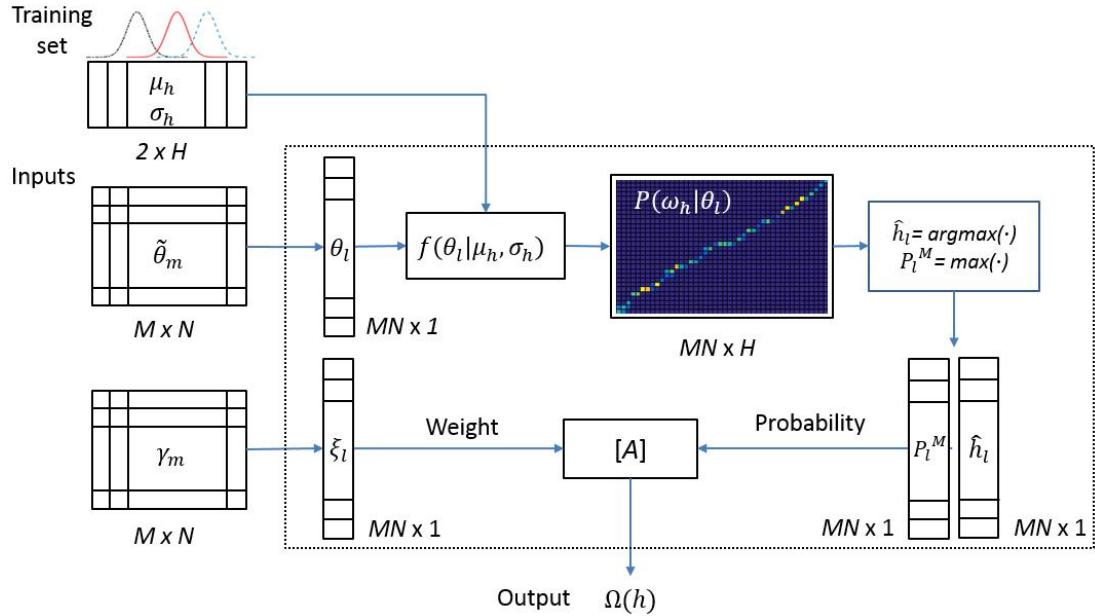
Figure 3.5. Block diagram of feature extraction process for harmonic order detector. The process [A] is described in Table 3.1

Table 3.1 Pseudo code for computing weighted probability of harmonics for [A] in Fig. 3.5

$$P^M(l) = \max \{P(\omega_h|\theta_l)\}$$

$$\hat{h}(l) = \arg \max_h \{P(\omega_h|\theta_l)\}$$

For $h = 1$ to H

    Find index i subject to $\hat{h}(l) = h$

    $C_h = \text{count}\{i\}$

    $\xi_j = \xi_l(i)$ and $P_j^M = P^M(i)$ where $j = 1, 2, \cdots$

    $\Omega(h) = \frac{1}{C_h} \sum_{j=1}^{C_h} \xi_j P_j^M$

End

the probability to minimize the noise peak effect. Then, the average of weighted probability is used as a feature for the harmonic order detector. Let us denote $\Omega(h)$ to be the weighted probability of harmonics (WPH) which is computed as Table. 3.1. Let θ_l and ξ_l shows 1-by-MN vector for all elements of $\tilde{\theta}_m$ and γ_m of N frames for the computation of $\Omega(h)$. More details about PDF estimation for $\tilde{\theta}_m$ and probability computation is described in following paragraphs.

3.2.2 Probability density

Although some of $\tilde{\theta}_m$ of HSP have common angles, each of them have different variation $\sigma_m{}^2$ due to different Rician factors. Therefore GMM is utilized to get the comprehensive pdf for that angles. Fig. 3.6 (a) and (b) shows PDF of θ_m and $\tilde{\theta}_m$ computed by simulation where $m = 1, \cdots, M$, $N = 10$ and $M = {}_N C_2$. The GMM for common reference angle $\tilde{\theta}_{ref_h}$ with different variation $\sigma_m{}^2$ is shown in Fig. 3.6 (c). GMM results shows good agreement with simulation results. Note that the normal distribution characteristics are preserved after conversion as shown in Fig. 3.7.
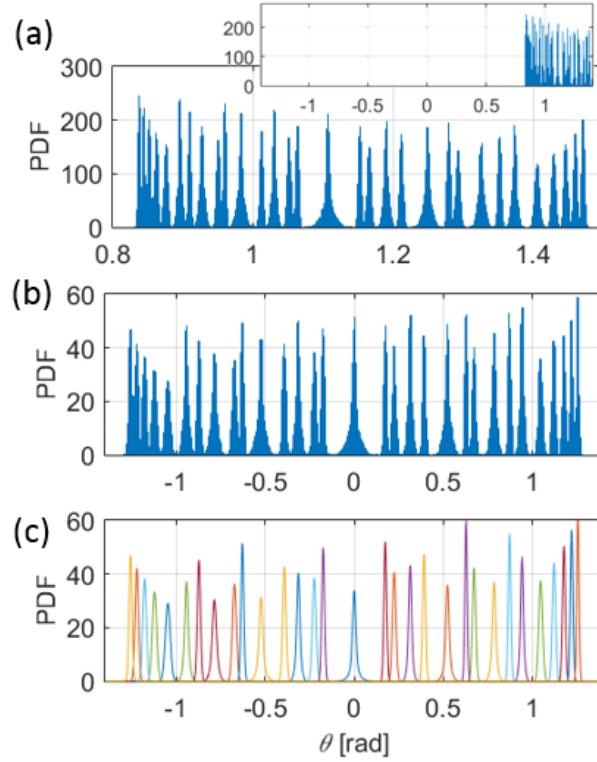
Figure 3.6. PDF of (a) θ_m and (b) $\tilde{\theta}_m$ by simulation for $f_0 = 50\,Hz$ and $\Delta f = 2\,Hz$ small window above (a) is view of $-\pi/2 \sim \pi/2$. (c) PDF estimation using GMM for $\tilde{\theta}_m$
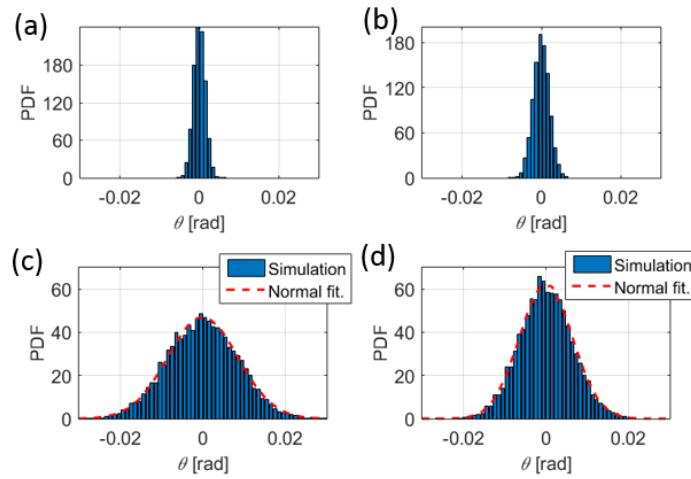
Figure 3.7. (a) Left end PDF and (b) right end PDF of θ_m. (c) Left end PDF and (b) right end PDF of $\tilde{\theta}_m$. Normal distribution characteristics are preserved after conversion.

3.2.3 Probability

Decision criterion for ω_h is as follow:

$$P(\omega_h|\tilde{\theta}_m) > P(\omega_{h'}|\tilde{\theta}_m) \tag{3.5}$$

where $h' = 1, \cdots, h-1, h+1, \cdots, H$. Given that prior probability $P(\omega_h)$ is the same with $P(\omega_{h'})$, $P(\omega_h|\tilde{\theta}_m)$ is regarded as likelihood function corresponding to PDF of $\tilde{\theta}_m$. Hence, they have different amplitude when $\tilde{\theta}_m = \tilde{\theta}_{ref_h}$ as shown in Fig. 3.6. In order to assign the same probability to each ω_h for $\tilde{\theta}_m = \tilde{\theta}_{ref_h}$, new variable y is introduced, which is defined as

$$y = -\left|\tilde{\theta}_m - \tilde{\theta}_{ref_h}\right|. \tag{3.6}$$

Cumulative density function (CDF) of eq. 3.6 is defined as

$$P(\omega_h|\tilde{\theta}_m) = 1 - \frac{1}{2}\left[\text{erf}\left(\frac{-\left|\tilde{\theta}_m - \tilde{\theta}_{ref_h}\right|}{\sigma\sqrt{2}}\right) + \text{erf}\left(\frac{-\left|\tilde{\theta}_m - \tilde{\theta}_{ref_h}\right|}{\sigma\sqrt{2}}\right)\right]. \tag{3.7}$$

Derivation of Eq. 3.7 from 3.6 is in the appendix B. Then CDF of y has 1 when $\tilde{\theta}_m$ is exactly matched with $\tilde{\theta}_{ref_h}$.

3.2.4 Classification

Extreme learning machine (ELM) is used for classifier.[21] It is single layer feed forward neural network classifier but gradient based back-propagation (BP) is not required, which is necessary for the training in convectional neural network. By calculating Moore-Pen-rose generalized inverse of the hidden layer output matrix with randomly selecting input weights and biases, ELM not only makes the learning fast but also produce good generalization performance.

Target labels for ELM training is presented in Table 3.2. One of output nodes corresponds with a one class for the multiclass classification in general so that the number one is assigned to the node while zeroes are assigned to the other nodes. However, each harmonic order is connected with the output node in this study. Hence, multiple activation of harmonic orders can be considered by assigning multiple ones into the output nodes. For example, if all harmonic orders are observed, all output nodes should be assigned by ones. If there are only two overtones are measured, two output node can be activated. In case of noise or non-harmonic spectrum, zeroes are assigned to all output nodes. Table 3.1 is proposed on the assumption that the harmonic orders are generated consecutively. In this study, 46 different cases are used to consider ten of harmonics.

Table 3.2 Target label of ELM training for harmonic order detection

Case	Number of harmonic orders									
	1	2	3	4	5	6	7	8	9	10
1	1	1	1	1	1	1	1	1	1	1
2	0	1	1	1	1	1	1	1	1	1
3	0	0	1	1	1	1	1	1	1	1
4	0	0	0	1	1	1	1	1	1	1
5	0	0	0	0	1	1	1	1	1	1
6	0	0	0	0	0	1	1	1	1	1
7	0	0	0	0	0	0	1	1	1	1
8	0	0	0	0	0	0	0	1	1	1
9	0	0	0	0	0	0	0	0	1	1
10	1	1	1	1	1	1	1	1	1	0
11	1	1	1	1	1	1	1	1	0	0
12	1	1	1	1	1	1	1	0	0	0
13	1	1	1	1	1	1	0	0	0	0
14	1	1	1	1	1	0	0	0	0	0
15	1	1	1	1	0	0	0	0	0	0
16	1	1	1	0	0	0	0	0	0	0
17	1	1	0	0	0	0	0	0	0	0
18	0	1	1	1	1	1	1	1	1	0
19	0	0	1	1	1	1	1	1	1	0
20	0	0	0	1	1	1	1	1	1	0
21	0	0	0	0	1	1	1	1	1	0
22	0	0	0	0	0	1	1	1	1	0
23	0	0	0	0	0	0	1	1	1	0
24	0	0	0	0	0	0	0	1	1	0
25	0	1	1	1	1	1	1	1	0	0
26	0	0	1	1	1	1	1	1	0	0
27	0	0	0	1	1	1	1	1	0	0
28	0	0	0	0	1	1	1	1	0	0
29	0	0	0	0	0	1	1	1	0	0
30	0	0	0	0	0	0	1	1	0	0
31	0	1	1	1	1	1	1	0	0	0
32	0	0	1	1	1	1	1	0	0	0

33	0	0	0	1	1	1	1	0	0	0
34	0	0	0	0	1	1	1	0	0	0
35	0	0	0	0	0	1	1	0	0	0
36	0	1	1	1	1	1	0	0	0	0
37	0	0	1	1	1	1	0	0	0	0
38	0	0	0	1	1	1	0	0	0	0
39	0	0	0	0	1	1	0	0	0	0
40	0	1	1	1	1	0	0	0	0	0
41	0	0	1	1	1	0	0	0	0	0
42	0	0	0	1	1	0	0	0	0	0
43	0	1	1	1	0	0	0	0	0	0
44	0	0	1	1	0	0	0	0	0	0
45	0	1	1	0	0	0	0	0	0	0
46	0	0	0	0	0	0	0	0	0	0

3.3 Results

3.3.1. Simulation data

Harmonic order detection is conducted using the proposed feature and simulated signals. Fig. 3.8 (a) and (b) are the simulated signals including time variation of the peak frequencies. In the Fig 3.8 (a), f_0 changes from 80 Hz to 140 Hz and six harmonic orders are observed from f_0 to 5th overtone. On the other hand, f_0 changes from 40 Hz to 110 Hz with six harmonics in Fig. 3.8 (b). However, f_0 and 1st overtone are missed and other six harmonics from 3rd to 8th overtones are observed. ELM results are shown in Fig. 3.8 (c) and 3.8 (d) for the simulated signals of Fig. 3.8 (a) and 3.8 (b), respectively. Even though peak frequencies change with time, the ELM results show the harmonic orders correctly and stably in both cases. Absence and variation of f_0 does not give any influence on the results as shown in Fig. 3.8 (d).
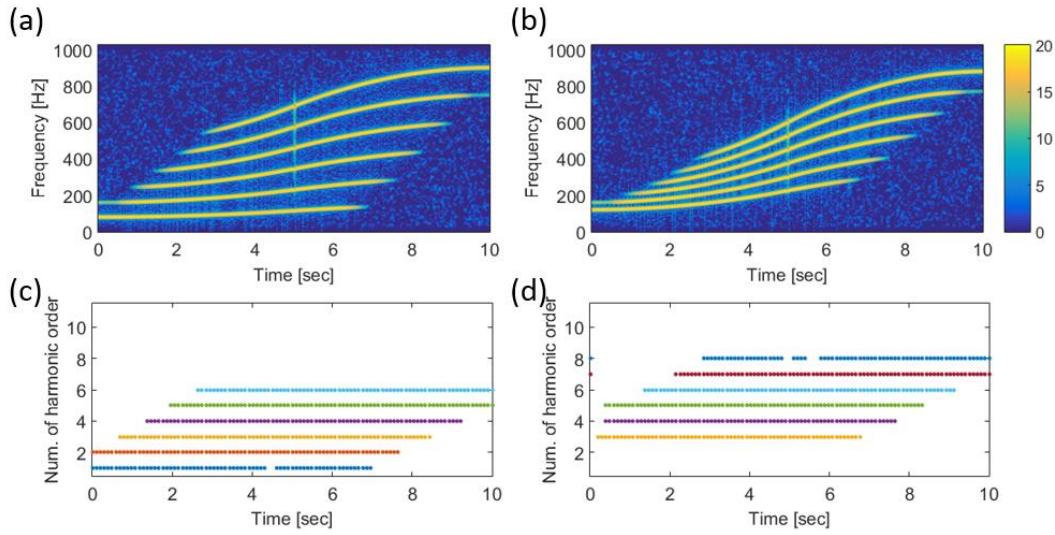
Figure 3.8. Simulated signals whose f_0 and harmonic order changes with time. (a) f_0 is from 80 Hz to 120 Hz and 6 harmonics are observed including f_0. (b) f_0 is from 80 Hz to 120 Hz and 6 harmonics are observed without f_0 and 1st overtone.

3.3.2 Measurement data

To evaluate the proposed feature in practical situation, measurement signals are used for harmonic order recognition. Fig. 3.9 (a) ~ 3.9 (d) show spectrograms of a UAV, drone, motorcycle, and car, respectively. ELM results are shown in Fig. 3.10 (a) ~ 3.10 (d). The ELM output of UAV signal correctly presents the change of harmonics with time. The harmonics disappear from 1 sec to 8 sec and from the 10th to 7th harmonic, as in Fig. 3.9 (a) and Fig. 3.10 (a). Fig. 3.9 (b) shows the sound of motorcycle. Three harmonics are recognized by ELM as shown in Fig. 3.10 (b). Since the sound quality is not very clear, 3rd harmonic is lost few time to time but it shows good agreement with original sound overall. Harmonic peaks are not clearly observed in drone sound because of many propellers as shown in Fig. 3.9 (c). Despite of complexity of harmonics, ELM results shows its harmonic order correctly as shown in Fig. 3.10 (c). Fig. shows 3.9 (d) represent car passing sound. There is no harmonic peaks and mostly is broad band spectrum. ELM results for the car passing sound is shown in Fig. 3.10 (d). No harmonic orders are identified as intended.
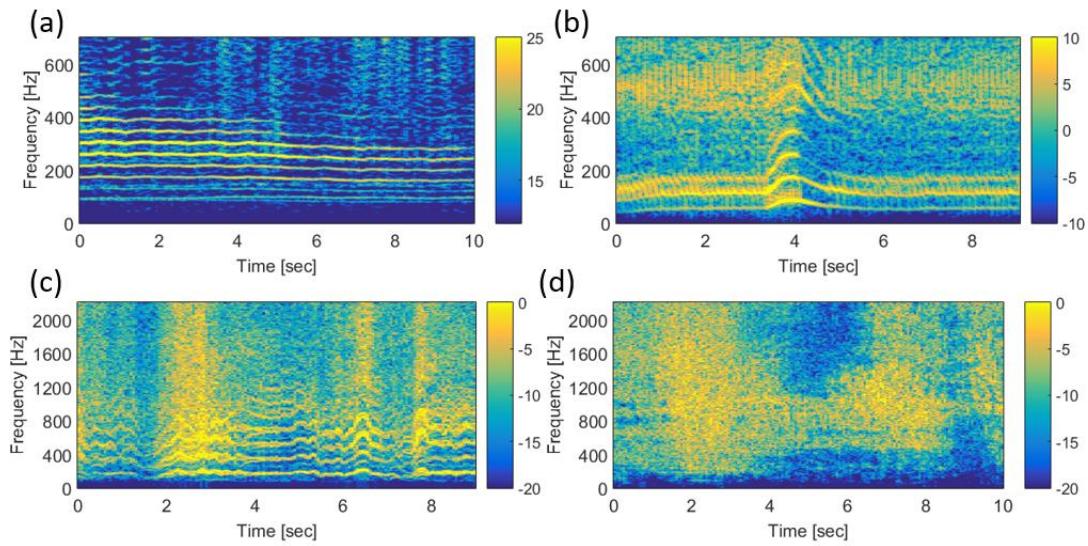
Figure 3.8. Measurement sound of (a) UAV, (b) motorcyle, (c) drone, and (d) car.
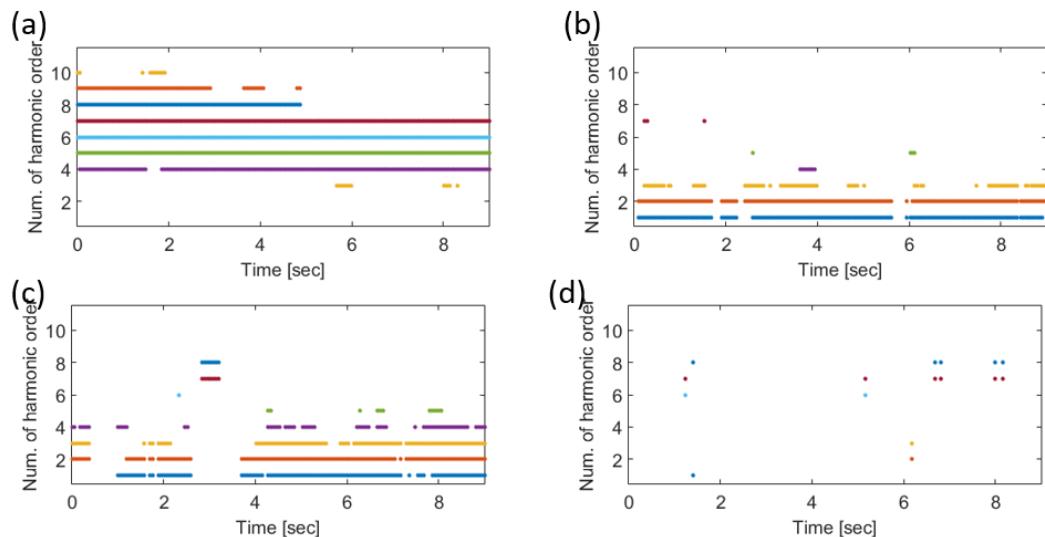
Figure 3.9. ELM results for the measurement sound of (a) UAV, (b) motorcyle, (c) drone, and (d) car.

Chapter 4

MULTICLASS CLASSIFICATION

4.1. Introduction

We have seen that WPH (Weighted probability of harmonics) can be used as an effective feature for harmonic order detection. In this chapter ELM output of the WPH is analyzed in detail and a new algorithm is proposed by using it for a multiclass classification. The new algorithm is inherently based on the concept of HLA. As mentioned in Chapter 1, HLA is composed of spectral magnitudes of harmonic orders. The crucial problem of HLA is to find the fundamental frequency by greedy search process. The ELM output of WPH exactly shows the relative magnitudes of the harmonic orders as HLA does. However, it does not require the fundamental frequency as we presented through the Chapter 3. Fig 4.1 (b) shows frequency spectrums and its mean spectrum of 1.5 second interval of harmonic signals in Fig. 4.1 (a). The spectral magnitude of 6th and 7th orders of harmonics are larger than the others and it decreases gradually for smaller orders while it decrease more sharply for higher orders. Fig. 4.1 (c) shows the mean ELM output of WPH with its variances. It comes out just before the decision step and harmonic orders can be noticed using a specific threshold value. It shows that the ELM output represents the relative magnitude of spectral peaks even without any process to find the fundamental frequency. Therefore, we can expect that the ELM output can be successful alternatives in HLA for multiclass classification for harmonic

signals.
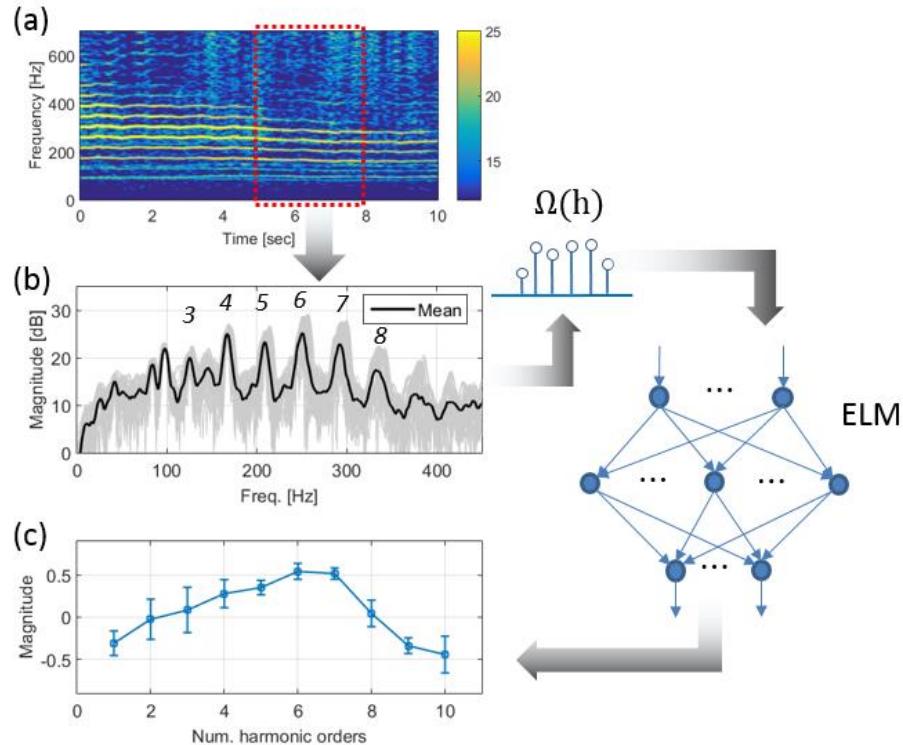
Figure 4.1. (a) An example of harmonic signal. (b) Frequency spectrums with mean spectrum of 1.5 second interval the example of harmonic signal. (c) Means and standard deviations of ELM output using the WPH extracted from the given frequency spectrums.

4.2. Algorithm

In order for the multiclass classification of harmonic signals, we propose a two stage ELM. It is composed of two consecutive ELM structure as shown in Fig. 4.2. The first ELM is the same as the harmonic order detector introduced in previous chapter but the output nodes are connected with the second ELM without final decision step using threshold value. Since the first ELM yields the unique characteristics of harmonic orders as described in the introduction, it takes a role of an additional feature extraction process and the second ELM is followed by the first ELM as a multiclass classifier. Since all weights of first ELM are trained by simulations, only second ELM need to be trained using real training dataset.

To investigate performance of the proposed algorithm, three classes of harmonic signals and one non-harmonic broad-band signal are used. Since the number of all classes are four, the number of output nodes is set to be four. The number of input nodes of the second ELM can be variable considering characteristics of harmonics. In this study, 10 input nodes are considered as the same value of the harmonic order detector in Chapter 3. Parameters for the second ELM is summarized in Table. 4.1.

For the comparison, single ELM classification using MFCC is used. Twenty Mel-filter banks and twelve cepstral coefficients are used for MFCC including log energy. Therefore, thirteen input nodes are used for ELM and the number of output nodes is the same as proposed one.
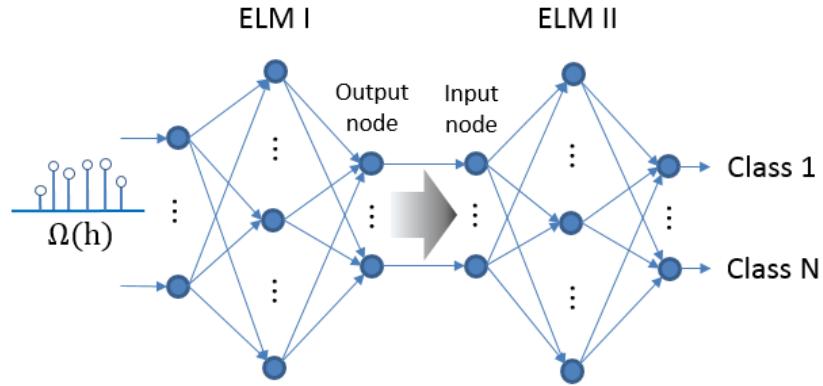
Figure 4.2. Two stage ELM for multiclass classification

Table 4.1. Parameters of the second ELM for four-class classification

Parameters	Values
Input node	10 (proposed)
	13 (MFCC)
Hidden node	100
Output node	4

4.3. Results

4.3.1. Experimental data

Three classes of harmonic signals and one non-harmonic broad-band signal is used for training and test. All data were downloaded or recorded from the websites. Fig. 4.3 shows representative experimental data of four classes which are used for training. Fig. 4.3 (a), (b), and (c) is sounds of UAV, drone, and motorcycle, respectively. Fig. 4.3 (d) is the sound of car passing in the road. Measured UAV sounds which are used in previous chapters are intentionally excluded in this study because their frequency spectrum shapes and magnitudes are totally different with drone and motorcycle signals. It yields trivial results so that we try to collect the sound data which have similar shape and level of spectrums with in frequency range. Collected data have 40 ~ 50 seconds of durations that is corresponding 640 ~ 800 frames. 400 frames are used for training and the others are used for test. Sampling frequency of original data was 44100 Hz but it was down sampled for 1/20 scale to make it becomes 2205 Hz of sampling frequency and band pass filtering is applied where its pass band is 50 ~ 2150 Hz.
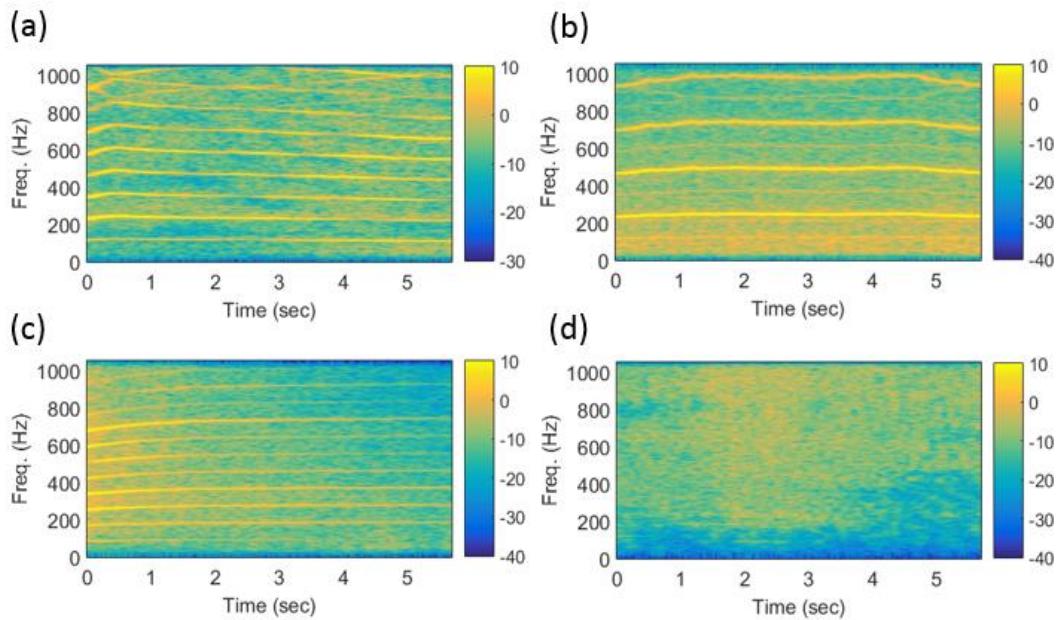
Figure 4.3. Sound of (a) UAV, (b) drone, (c) motorcycle, and (d) car passing

4.3.2. Results

The results of multiclass classification are shown in Fig. 4.4. Fig. 4.4 (a) shows the results of typical ELM with MFCC. The UAV and motorcycle sound are correctly classified but there are some errors for drone and car sound. Overall recognition error rate is 22.9 %. Most of errors are from drone sound. Causes of the error is analyzed in detail in 4.3.3. Recognition error rate is 6.7 % for proposed algorithm. Most of data are correctly classified by the proposed one.
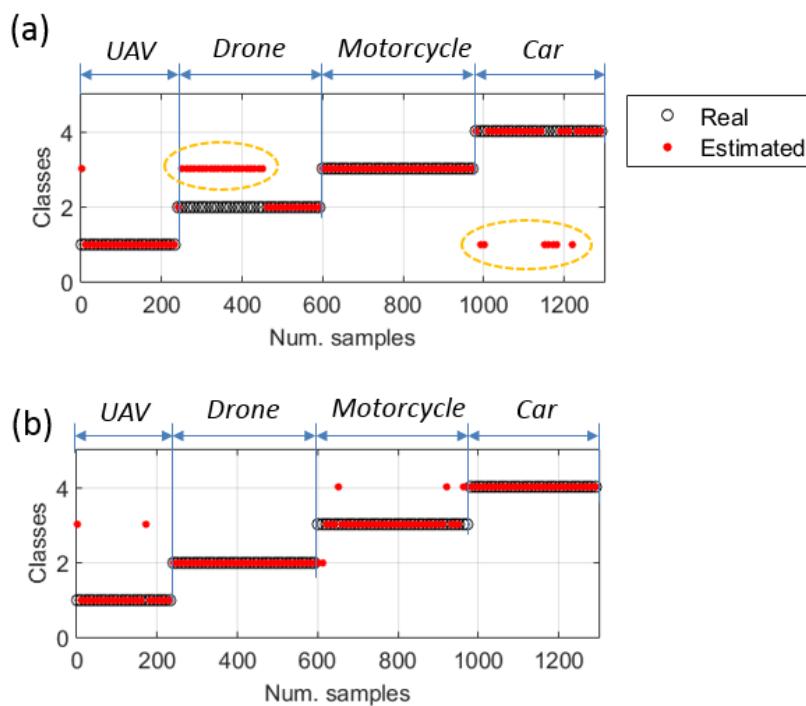
Figure 4.4. (a) Classification result by typical ELM structure MFCC and (b) by the proposed two stage ELM with WPH

4.3.3. Analysis

The fundamental frequency of drone sound used for training is nearly 250 Hz and it lasted as shown in Fig. 4.5 (a). However, the fundamental frequency and its overtones of test set are changed with time and some of them are 40 Hz lower than that of training sound as shown in Fig. 4.5 (b). According to the classification results in Fig. 4.4 (a), errors occur in 0.5 ~ 7 seconds where the fundamental frequency and its overtones are lower than its training data.

Fig. 4.6 (a) shows that the Mel-filter bank energies of training and test data. Filter bank energies are shifted by the frequency variation. The first peaks are located in 7th filter but it is
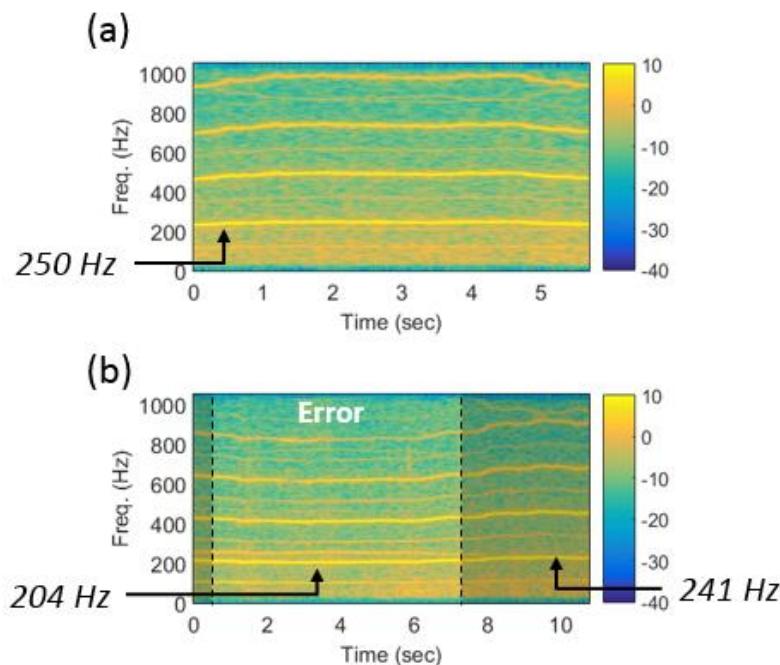
Figure 4.5. (a) A part of training data of drone, (b) A part of test data of drone

moved to 6th. Variations of filter bank energies by overtones are more significant. The shift of filter bank energy is gives rise to the phase difference of MFCC between training data and test data as shown in Fig. 4.6 (b). The MFCC of test data of drone sound is rather similar to that of motorcycle so that it leads to the error in classification results.

On the other hand, even though the fundamental frequency and its overtones of test data differ with training data, the first ELM output of proposed algorithm yields more stable
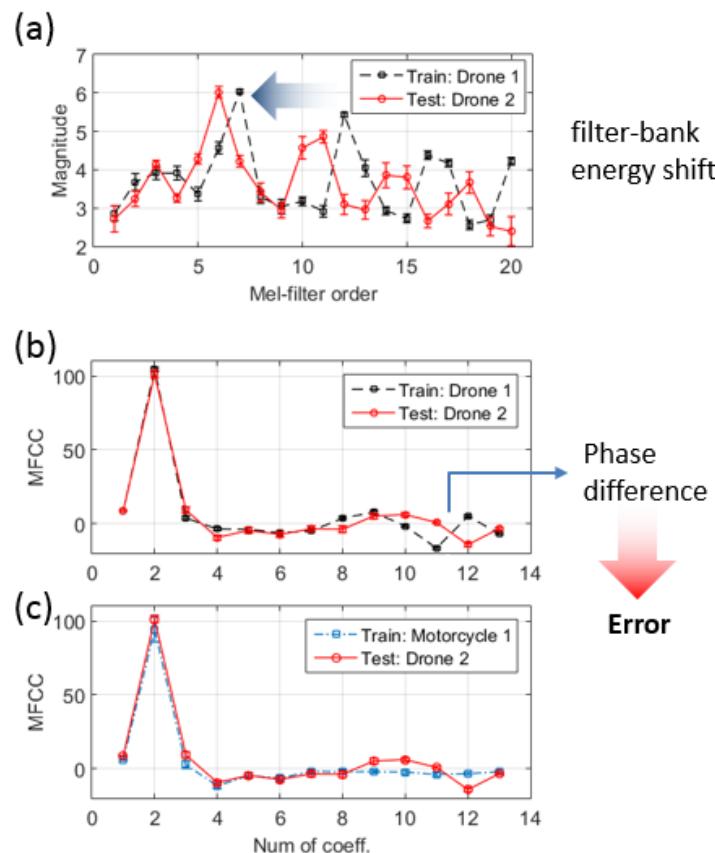
Figure 4.6. (a) Filter bank energy of training data and test data of drone sound. (b) MFCC of training data and test data. (c) MFCC of training data of motorcycle and test data of drone.

features as shown in Fig. 4.7 (a). First four orders of harmonics are get higher values in both

training and test data. It has unique values and shapes comparing the other classes as shown

in Fig. 4.7 (b). The test drone signal was confused by the motorcycle training data for MFCC

but they are obviously different in case of proposed method. 2 ~ 4th order of harmonics and

8th order of harmonics are dominant. It follows the relative amplitude of spectral peaks as

shown in Fig 4.7 (c).
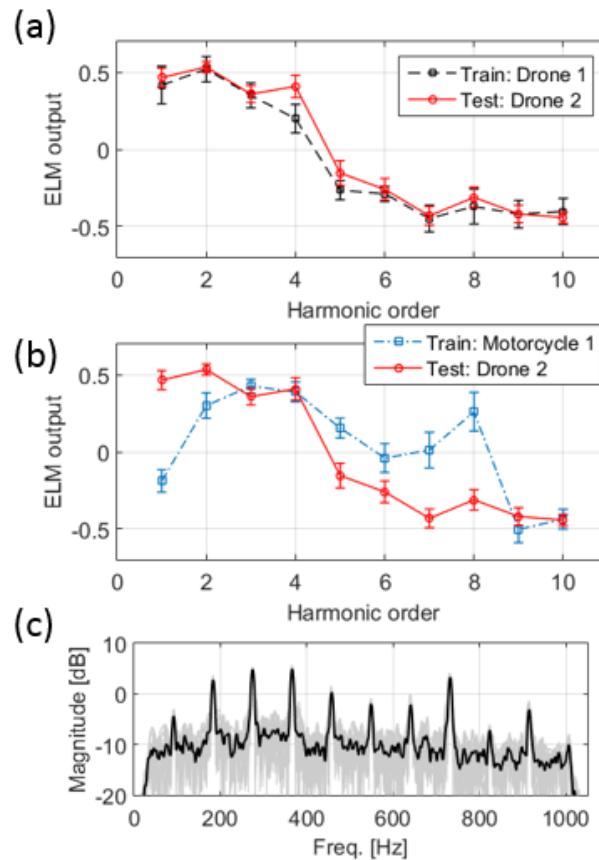
Figure 4.7. (a) First ELM outputs of training and test drone data. (b) First ELM output of

trained data of motorcycle and test data of drone (c) spectrums and the mean spectrum of

motorcycle

Chapter 5

CONCLUSIONS

Sound detection and classification algorithm based on harmonic spectral peaks were studied in this thesis. Firstly, binary classification algorithm was presented for harmonic sound and noise. By adopting two dimensional vector representation of spectral peaks and evaluating similarity between the input vectors with the reference vectors, harmonic spectrum could be identified from the other non-harmonic sound. For the similarity computation, two algorithm was introduced: inner product based algorithm and virtual array algorithm. The Gaussian mixture PDF of inner product values computed by normalized input vectors and reference vectors was utilized for the training process of the former algorithm. It does not require complex processes but the measurement harmonic sound was successfully classified from the noise. Angles of the vectors were used for similarity function in the latter algorithm. By adopting array signal processing, the angle of input vectors could be compared with the reference angle. By evaluating detection probability using ROC and UAC comparing with MFCC, we shows that the proposed feature exhibits robust and stable performance.

Secondly, vector representation was expended to harmonic order detection. By using PDF of all orders of harmonics and ELM, harmonic order information of unknown input signals could be estimated. We proposed WPH as a new feature vector to detect the order of harmonics. Its performance was proved by ELM classifier using both simulation data and

measurement data.

Lastly, multiclass classification was performed and analyzed. Based on WPH, the two stage ELM structure was suggested. The multiple classes of harmonic sound could be successfully identified by proposed algorithm while it was partially failed by MFCC. This is based on the key idea of HLA which is useful and effective feature with using a set of spectral magnitude of harmonic orders. Even though HLA has crucial drawback of fundamental frequency dependence with greedy search process, it has presented that the proposed algorithm can overcome the limitations.

Since the measurement data is required to train a classifier weights, unlike the binary classification which is used simulation data, sufficient data are needed. However, small amount of data is used in this thesis so that it is limited to generalize from this results. Therefore, more analytical study using sufficient data set is planned for the future work. Note that we present the potential for multiclass classification of harmonic signals in more practical way. Proposed algorithm can be applicable not only the vehicle sounds classification but also the other fields such as bubble cavitation detection in medical ultrasound area or the hornet sound recognition to protect honeybee from hornet's attack for apiculture. Proposed algorithm can be widely used because harmonic sounds can be observed in many area of nature.

APPENDIX A

The scalar product of two unit vectors in eq. (6) can be expressed by cosine of angle difference $d\theta$, where $d\theta = \theta_{ref} - \theta_m$. Since a θ_m which yields w_M shows small $d\theta$ usually, small angle approximation can be applied into scalar product as bellow

$$w = \cos^2 d\theta = \left(1 - \frac{1}{2}d\theta^2 + \cdots\right)^2 \cong 1 - d\theta^2 \qquad (A.1)$$

Signal: According to the Rician fading theory in wireless communication, it is well known that θ_m follows normal distribution when both f_i and f_j are normal distribution by virtue of the large Rician factor k defined as $k = \left(f_i^2 + f_j^2\right)/2\sigma_f$.[23] Then, θ_m can be written as follows normal distribution whose mean is θ_{ref} and standard deviation is σ_k, where $\sigma_k = 1/\sqrt{2k}$. Therefore, it is obvious that $d\theta$ has normal distribution with zero mean.

$$f_{d\theta}(d\theta) \sim N(0, \sigma_k) \qquad (A.2)$$

By Eq. (A1), cumulative distribution function (CDF) of w can be expressed as

$$F_w(w) = P(1 - d\theta^2 \leq w) = P\left(|d\theta| \geq \sqrt{1-w}\right)$$

$$= \begin{cases} F_{d\theta}\left(-\sqrt{1-w}\right), & d\theta < 0 \\ 1 - F_{d\theta}\left(\sqrt{1-w}\right), & d\theta \geq 0 \end{cases}. \qquad (A.3)$$

Due to the symmetry of $f_{d\theta}$, we can rewrite eq (A.3) as follow

$$F_w(w) = 2F_{d\theta}\left(-\sqrt{1-w}\right) \qquad (A.4)$$

By the hypothesis of Eq. (6), r_m of target signal yields the largest w so that we can regard $F_w(w) = F_{w_M}(w)$. Hence, we conclude that PDF of w_M is written as follow

$$f_{w_M}(w_M) = \frac{1}{\sqrt{2\pi{\sigma_k}^2}}(1-w_M)^{-\frac{1}{2}}e^{-\frac{1-w_M}{2{\sigma_k}^2}}$$

$$= \frac{(1-w_M)^{-\frac{1}{2}}e^{-\frac{1-w_M}{(2{\sigma_k}^2)}}}{(2{\sigma_k}^2)^{\frac{1}{2}}\Gamma\left(\frac{1}{2}\right)} = \text{gamma}(1-w_M, \alpha, \beta) \tag{A.5}$$

where $\alpha = 1/2$ and $\beta = 2{\sigma_k}^2$.

Noise: Since f_i and f_j are uniform distribution subjected to $f_j > f_i$, their joint PDF is expressed as

$$f_{f_i,f_j}(f_i, f_j) = 2/{f_s}^2 \tag{A.6}$$

When we apply coordinate transformation from Cartesian to polar considering $l_m = \sqrt{{f_i}^2 + {f_j}^2}$, $\theta_m = \tan^{-1}\left(f_j/f_i\right)$ and $J(l_m, \theta_m) = l_m$, where J represent Jacobian. Then joint PDF in polar coordinate is written as follow:[24]

$$f_{l_m,\theta_m}(l_m, \theta_m) = l_m f_{f_i,f_j}(f_i, f_j) = 2l_m/{f_s}^2, \tag{A.7}$$

where $\pi/4 \leq \theta_m \leq \pi/2$ and $0 \leq l_m \leq \sqrt{(f_s/2)^2 + {f_i}^2} = (f_s/2)\csc\theta$. Then the marginal distribution of with respect θ_m to is expressed by $\csc^2\theta_m$. Therefore, PDF of $d\theta$ is defined as

$$f_{d\theta}(d\theta) = \csc^2(\theta_{ref} - \theta_m), \tag{A.8}$$

which is not symmetry for $d\theta$. When we apply Eq. (A.8) into Eq. (A.3), PDF of w is yield as follow:

$$f_w(w) = \begin{cases} \dfrac{1}{2\sqrt{1-w}}\operatorname{cosec}^2\big(\theta_{ref} - \sqrt{1-w}\big), a \le w \le b \\ 1 + \dfrac{1}{2\sqrt{1-w}}\big\{\operatorname{cosec}^2\big(\theta_{ref} + \sqrt{1-w}\big) \\ \qquad + \operatorname{cosec}^2\big(\theta_{ref} - \sqrt{1-w}\big)\big\}, b < w \le 1 \end{cases} \tag{A.9}$$

where $a = 1 - \big(\theta_{ref} - \pi/2\big)^2$ and $b = 1 - \big(\theta_{ref} - \pi/4\big)^2$

According to extreme value theory, maximum value of $f_w(w)$ follows Weibull distribution which belongs to exponential family as gamma distribution does.[25] Therefore we treat $f_{w_M}(w)$ as gamma distribution using gamma parameter estimation.

APPENDIX B

As described in Appendix A, the frequency vector angles of harmonic spectral peaks follows normal distribution. We assume that the converted angles are also have normal distribution because variation of the normal distribution has small as followed:

$$\tilde{\theta}_m \sim N(\tilde{\theta}_{ref_h}, \sigma_h), \tag{B.1}$$

where $\tilde{\theta}_{ref_h}$ is a unique reference angle and σ_h is the standard deviation. Another variable Y is adopted to measure a distance of $\tilde{\theta}_m$ and $\tilde{\theta}_{ref_h}$. It is defined as

$$Y = -\left|\tilde{\theta}_m - \tilde{\theta}_{ref_h}\right|. \tag{B.2}$$

Then, the cumulative distribution function (CDF) of Y can be computed as followed:

$$F_Y(y) = P(Y \leq y) = P(-\left|\tilde{\theta}_m - \tilde{\theta}_{ref_h}\right| \leq y)$$

$$= P(\left|\tilde{\theta}_m - \tilde{\theta}_{ref_h}\right| \geq -y) \tag{B.3}$$

$$= 1 - F_{\left|\tilde{\theta}_m - \tilde{\theta}_{ref_h}\right|}(-y)$$

Since $\tilde{\theta}_m - \tilde{\theta}_{ref_h}$ follows $N(0, \sigma_h)$, its absolute values are defined as

$$\left|\tilde{\theta}_m - \tilde{\theta}_{ref_h}\right| \sim Folded\ Normal \tag{B.4}$$

The CDF of folded normal distribution is given by

$$F_Y(y) = 1 - \frac{1}{2}\left[\text{erf}\left(\frac{-y}{\sigma\sqrt{2}}\right) + \text{erf}\left(\frac{-y}{\sigma\sqrt{2}}\right)\right] \tag{B.5}$$

Finally, when we regard the probability of ω_h given $\tilde{\theta}_m$ which is defined as $P(\omega_h|\tilde{\theta}_m)$ in

Chapter 3 as the CDF of Y of eq. (B.2), it is defined as

$$P\big(\omega_h\big|\tilde{\theta}_m\big) = 1 - \frac{1}{2}\left[\text{erf}\left(\frac{-\big|\tilde{\theta}_m - \tilde{\theta}_{ref_h}\big|}{\sigma\sqrt{2}}\right) + \text{erf}\left(\frac{-\big|\tilde{\theta}_m - \tilde{\theta}_{ref_h}\big|}{\sigma\sqrt{2}}\right)\right] \tag{B.6}$$

BIBLIOGRAPHY

[1] J. R. Wilson, "UAV roundup 2013," Aerospace America, 51(7), 26-36 (2013).

[2] INEA conulting, "Global commercial and civil uav market guide 2014-2015," http://renewablemarketwatch.com/country-reports

[3] R. E. Weibel and R. J. Hansman, "Safety considerations for operation of different classes of UAVs in the NAS," *AIAA 4th Aviation Tehcnology, Integration and Operations Forum, AIAA 3rd Unmanned Unlimited Technical Conference, Workshop and Exhibit.* 1-11, (2004).

[4] E. B. Carr, "Unmanned Aerial Vehicles: Examining the Safety, Security, Privacy and Regulatory Issues of Integration into US Airspace," National Centre for Policy Analysis (NCPA). Retrieved on September 23 (2013): 2014.

[5] E. E. Case, A. M. Zelnio, and B. D. Rigling, "Low-cost acoustic array for small UAV detection and tracking," in *Proceeding of Aerospace and Electronics Conference, NAECON, IEEE National*, 110-113 (2008).

[6] D. Miljković, M. Maletić, and M. Obad, "Comparative Investigation of Aircraft Interior Noise Properties," *In 3rd Congress of the Alps Adria Acoustics Association*, (2007).

[7] D. Miljković, J. Ivošević, and T. Bucak, "Two vs. Three Blade Propeller-Cockpit Noise Comparison," *In Proceedings 5th Congress of Alps-Adria Acoustics Association*, 12-14, (2012).

[8] G. Sinibaldi and L. Marino, "Experimental analysis on the noise of propellers for small UAV," Applied Acoustics, 74(1), 79-88 (2013)

[9] T. Pham, N. Srour, "TTCP AG-6: acoustic detection and tracking of UAVs," in *Proceeding. SPIE 5417, Unattended/Unmanned Ground, Ocean, and Air Sensor Technologies and Applications VI,* 24-30 (2004).

[10] M. Wellman and N. Srour, "Enhanced target identification using higher order shape statistics," *(No. ARL-TR-1723) Army research lab adelphi MD sensors and electron devices directorate*, (1999).

[11] W. Shi, B. Bishop, G. Arabadjis, J. Yoder, P. Hill, and R. Plasse, "Detecting, Tracking, and Identifying Airborne Threats with Netted Sensor Fence," INTECH Open Access Publisher, 139-158, (2011).

[12] H. L. Van Trees, "Detection, estimation, and modulation theory, optimum array processing," John Wiley & Sons, (2004).

[13] S. S. Haykin and A. O. Steinhardt. "Adaptive radar detection and estimation," Vol. 11. Wiley-Interscience, (1992).

[14] K. P. Murphy, "Machine learning: a probabilistic perspective," MIT press, (2012).

[15] S. M. Kay, "Fundamentals of statistical signal processing, Vol. II: Detection Theory," Signal Processing. Upper Saddle River, NJ: Prentice Hall (1998).

[16] J. L. Melsa, and D. L. Cohn, "Decision and estimation theory," McGraw-Hill, (1978).

[17] G. G. Roussas, "A course in mathematical statistics," Academic Press, (1997).

[18] Fawcett, Tom. "An introduction to ROC analysis," *Pattern recognition letters*, 27(8), 861-874 (2006).

[19] D. G. Childers, D. P. Skinner, and R. C. Kemerait, "The cepstrum: A guide to processing." *Proceedings of the IEEE* 65(10), 1428-1443 (1977).

[20] Y. Qi and R. E. Hillman., "Temporal and spectral estimations of harmonics-to-noise ratio in human voice signals." *The Journal of the Acoustical Society of America*, 102(1), 537-543 (1997).

[21] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme Learning Machine: Theory and Applications", *Neurocomputing*, 70, 489-501, (2006).

[22] T. Piboongungon, V.A. Aalo, C.D. Iskander, and G.P. Efthymoglou, "Bivariate generalised gamma distribution with arbitrary fading parameters," *Electronics Letters*, 41(12), (2005).

[23] J. K. Cavers, "Mobile Channel Characteristics," Shady Island Press, (2004).

[24] A. Papulis and S. U. Pillaim "Probability, random variables and stocahstics processes," fourth edition, McGraw-Hill (2002).

[25] H. Rinne, "The Weibull distribution: a handbook," CRC Press, (2008).