Gender Identification based on Voice Signal Characteristics

ShivamChaudhary
Electronics & Communication Engineering
Meerut Institute of Engineering & Technology
Meerut
(Email: shivamchaudhary1001@gmail.com)

Davendra Kumar Sharma
Electronics & Communication Engineering
Meerut Institute of Engineering & Technology
Meerut
(Email: d_k_s1970@yahoo.co.in)

***Abstract*- In the present scenario gender identification is based on voice signal of human being. Several Automatic speech recognition systems uses gender identification and has proved to be of great importance. In today's technology gender identification is for speaker's identity in advance security system. In the proposed work, gender identification is done from voice signal by extracting the characteristics such as pitch, energy and mfcc. The features of voice signal is being classified using SVM classifier. Data base includes 280 speech files of which 140 samples are males and 140 samples are females. Training and testing is done on 80% and 20% of data base respectively. The classifier accuracy obtained is 96.45%.**

***Keywords*: gender identification, energy, pitch , MFCC, SVM classifier.**

I. INTRODUCTION

Voice is the most common form to convey information. Many things can be find out from voice such as sequence of words, gender, emotion, age and dialect. Gender identification is method that motives to determine the sex of the person through his voice signal characteristics. Automatically detecting the gender of a person has many applications from the point of view of automatic speech recognition, gender dependent models are more accurate than gender independent. Automatically detection of the gender of a speaker has many applications in many fields such as sorting telephone call by gender in automatic speech recognition system to enhance adaptability. Automatic speech recognition system which has gender specific models result in higher recognition rates as compared to gender independent one s. Speech coders which are gender specific with gender classifier are more accurate than otherwise. There are numerous applications of gender identification such as semantic information from multimedia, automatic answering machine, machine dialogue system for gender classification and others.

Researchers have posed various approaches for gender identification. To represent speech signal, most commonly used features are energy, pitch, frequency, formant frequency some general features such as autocorrelation coefficients, linear prediction coefficients and MFCC. SVM, GMM, ANN, HMM are used as classifiers.

II. RELATED WORKS

E.S.Paris *et al.* [1] proposed a technique for gender identification combining acoustics and pitch analysis. For acoustics analysis HMM and LDA is used for pitch estimation IMBE speech coding is used .

S. Slomka *et al.* [2] proposed fusion of multiple knowledge sources using a linear classifier. In this multiple knowledge sources uses parameterisation (mel base cepstral coefficients, reflection coefficients, log area ratios) then the paired GMM trained and tested with parametrisation are used and then the linear classifier is used.

R.D. R Fayunder *et al.* [3] studied automatic gender identification by voice signal using eigen filtering based on hebbian learning. In this artificial neural network a maximum eigen filter is implemented and trained via generalised hebbian learning The principle component analysis is employed to get the maximum information from voice signal. This technique is a effective way to show voice characteristics .

H. Harb *et al.* [4] proposed gender identification using a general audio classifier. This technique has robustness to adverse audio compression. In general audio classifier potential feature that can be used to obtain feature vector as MFCC features. Classifier which is used is neural network classifier.

A.Lindgren *et al.* [5] proposed speech recognition using reconstructed phase space features. Generation of a processing space called a reconstructed phase is done by using the theoretical results derived from non linear dynamics. Isolated phoneme experiments were performed to discover the effectiveness of these features.

H. Harb *et al.* [6] studied application of voice based gender identification in multimedia. In this, they showed fusion of features with different classifier.

F.Yingle *et al.* [7] proposed speaker gender identification based on combining non-linear and linear features. Linear parameter such as fractal dimension and fractal complexity are used. Firstly using the lifting method to get the pitch and then the exatraction of the speech signal fractal dimension. Finally as per the taken's theorem, reconstruction of phase space if fractal dimension sequence is done using time delay method, approximate entropy is calculated to find fractal dimension complexity.

M.Abdollahi *et al.* [8] proposed voice based gender identification via adaptive multiresolution classification of spectro temporal maps

M. A.Keyvanrad *et al.* [9] proposed a two layer classification fusion technique for AGI.

R.Djemili *et al.* [10] proposed a gender identification system using four classifier. Four different classifiers are used in the task of automatic speech based gender identification.. Four classifiers used are GMM, MLP, VQ and learning vector quantization along withmel frequency coefficients.

Ghosal.A *et al.* [11] proposed automatic male female voice discrimination. In this speech signal time domain features such as zero crossing rate, short time energy and spectral flux which is a frequency domain feature. Training and testing carried out using RANSAC and neural-net classifier.

G.S.Archana *et al.* [12] proposed gender identification and performance analysis of speech signals. Pitch is generally used for gender identification in male and female voice. Results showed that svm classification results were better than artificial neural network results.

Kumari.M *et al.* [13] proposed a new gender detection algorithm considering the non-stationarity of speech signal. In this speech signal feature pitch is found. Using pitch gender is determined. If it is high it is female, if low it is male. Comparision for high and low done on threshold basis.

Ramdinmawii.E *et al.* [14] proposed gender identification from speech signal by examining the speech signal production characteristics in this speech signal features such as pitch, mel frequency coefficients, signal energy are find out. Training and testing of data was done using SVM classifier with mel frequency coefficients accuracy was found to be 69%.

Shan. Li *et al.* [15] proposed the multidimensional speaker information recognition system, features such as pitch, formant,energy, linear prediction cepstrum coefficients(LPCC) and Mel-frequency cepstrum coefficients(MFCC) are concatenated to work as feature vectors and for classification SVM classifier is used.

III. ANALYSIS OF FEATURES OF SPEECH SIGNAL

There is physiological difference in the speech signal of male and female. Owing to these physiological differences there are some speech signal characteristics which changes in male and female voice. These changes helps us to differenciate gender on the basis of speech signal characteristics. Characteristics such as Energy, Pitch, MFCC etc. Finding the values of these features helps us to determine gender and classifiers such as SVM are used to classify the data based on these features.

A. Energy

The energy of a discrete time signal is defined as

$$E=\sum_{n=-\infty}^{+\infty} x^2(n) \qquad (1)$$

The above expression has little meaning for speech signal as it give small information about the time dependent properties of speech signal. To have more information about speech signal, short time energy is calculated. Commonly speech signal have regions such as voiced, unvoiced, silence and noise regions.Short time energy is given as

$$E_n=\sum_{n=m-N-1}^{m} x^2(n) \qquad (2)$$

Short time energy at sample n is simply the sum of energies of the N samples. Short time energy can also be defined as

$$E_n=\sum_{m=-\infty}^{\infty} [x(m)w(n-m)]^2 \qquad (3)$$

Where E_n is the energy at sample n in the signal x, w is the window and m is the number of frames occurred in the signal.

B. Pitch

Pitch is an important parameter revealing speaker's identity. It is considered only during the voiced segments of the speech signal. During unvoiced segments where there is no glottal vibrations, pitch is not considered or it is undefined. There are various methods which are available for estimation of pitch, such as autocorrelation method which is a time domain method , and cepstrum method which is a frequency domain method. Method used for estimation of pitch in this work is cepstrum method. Steps involved in the cepstral method are shown in fig. 1.

1) Take input speech signal.
2) Multiply the speech signal with the short duration window to get the frame.
$S(n), \quad 0 \leq n \leq N-1$
3) Take the DFT of S(n) to get S(k) and then take its magnitude |S(k)|.
4) Now take the IDFT of S(k) to set the cepstral C(n).

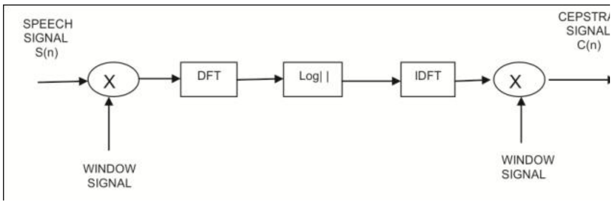

Fig. 1 Block diagram for pitch estimation using cepstrum

C.MFCC

MFCC is Mel frequency cepstralcoefficiens. MFCC takes into consideration the human perception sensitivity with respect to frequencies so it is best feature for speaker recognition. Steps involved in MFCCcomputation are shown in fig. 2.

1) Pre-Emphasis: The speech signal S(n) is given to a high pass filter. The aim of pre-emphasis is to boost the high frequency components.

$$Z(n)=Z(n)-\propto Z(n-1) \qquad (4)$$

Where ∝ is called as pre-emphasis coefficient, ranging between 0.9 to 1.0

2) Frame Blocking: The input speech signal is segmented into frames of 25ms (standard size) with overlap of 15 ms. Sampling rate is assumed at16,000hz (0.025*16000=4000 samples), frame shift taken is 10ms which allowed some overlap of frames. The first sample started at sample 0, the next 400 sample frame starst at sample 160. This process continues until it get end of the speech file..

3) Hamming Window: Now the next step is the windowing of each individual frame for the minimization of signal discontinuities at the beginning and end of each frame. For this, each frame has to be multiplied with a hamming window (W(n)). Signal in the frame is denoted by S(n), n=0,------N-1.

$$W(n,a)= (1-a)-a\cos(\frac{2\pi n}{N-1}) \qquad \text{where } 0\leq n\leq N-1 \ (5)$$

4) Fast Fourier Transform(FFT): FFT converts frame of N samples from time domain to frequency domain using the formula given below

$$z_i(k)=\sum_{n=1}^{N} S_i(n)h(n)e^{\frac{-2\pi}{N}} \qquad (6)$$

Where $S_i(n)$ is the signal in time domain and $S_i(k)$ is the signal in frequency domain from 1 to K, h(n) is the window with N samples long and K is the length of the FFT.

*5) Triangular Band Pass Filter (Mel Filter Bank):*Magnitude frequency response obtained is multiplied by a set of 20 triangular band pass filters and the log energy of each triangular band pass filter is obtained. These filters are equally placed along the melfrequency . Relation between linear frequency and mel frequency is given by following equation

$$Mel(f)=1125\ln(1+\frac{f}{700}) \qquad (7)$$

Mel frequency is directly proportional to the logarithm of the linear frequency.

6) Discrete Cosine Transform(DCT) : In this step DCT is applied to the log energy obtained from the triangular filter banks to have L mel scale cepstral coefficient expression for DCT is given as

$$C_m=\sum_{k=1}^{N} \cos(m(k-0.5)\frac{\pi}{N})E_k \qquad (8)$$

Where m=1,2------L, L=12

N=Number of triangular filter banks. N=20

E_k=Log energy obtained from 20 triangular band pass filters.

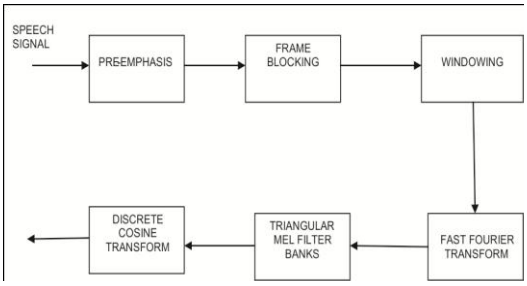

Fig. 2 Block diagram showing the steps in MFCC computation.

D. Classifier

The classification for speech signals is carried out using the Support Vector Machine(SVM). The primary focus in SVM is to adjust the the decision boundry that distinguishes different class labels. Figure 3 depicts the separation of two classes with SVM. The data points which are nearer to the hyperplane are called support vectors.

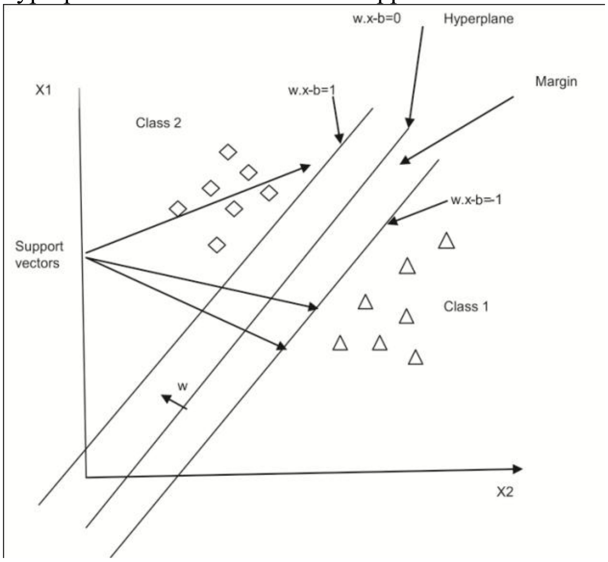

Fig 3 Separation of groups using SVM

IV. PROPOSED ALGORITHM

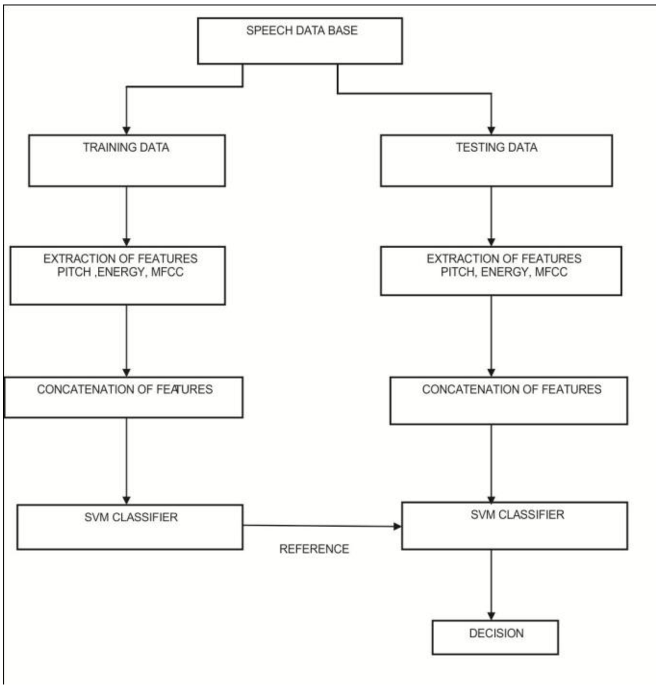
SPEECH DATA BASE
TRAINING DATA
TESTING DATA
EXTRACTION OF FEATURES PITCH ,ENERGY, MFCC
EXTRACTION OF FEATURES PITCH, ENERGY, MFCC
CONCATENATION OF FEATURES
CONCATENATION OF FEATURES
SVM CLASSIFIER
SVM CLASSIFIER
REFERENCE
DECISION

Fig. 4 Flowchart

The proposed algorithm identifies gender from speech signal using three different characteristics and a classifier

Speech data base is collected from TIMIT speech corpus, which has 140 speech files of males and 140 speech files from females. 80% of samples are used for training and 20% are used for testing. Speech signal features such as pitch, energy and MFCC are extracted from speech signal and concatenated. SVM classifier uses these concatenated features for the classification of data. Decision is made by the SVM based on its training on data. Steps are given below.

1) Speech data base is collected.
2) 80% is training data and 20% is testing data.
3) Speech characteristics such as pitch, energy, mfcc are extracted from speech signal.
4) Concatenation of these features is done.
5)SVM classifier is used for classification of data.
6) Sample is tested using reference with SVM classifier.
7) Decision is given whether speaker is male or female.

V. DATABASE

Data used for the implementation of this work is TIMIT data base. TIMIT is Texas Instruments Massachusetts Institute of Technology. It consists of 6300 sentences in total in which 10 sentences were spoken by each of 630 speakers of the United States. In this work we have taken speech files from 140 male speakers and 140 female speakers. The sampling frequency used is 16kHz.

VI. SIMULATION AND RESULTS

TIMIT speech corpus data based is used for collection of speech files. 140 male speech files and 140 female speech files are collected. 80% of the speech files are used for training and 20 % of speech files are used for testing. To obtain the results, MATLAB software is used for Simulation.

A. Comparison of pitch and energy

Fig. 5 and fig. 6 shows the Energy plot and Pitch plot for female and male speakers. By comparing Fig. 5 and fig. 6, it can be seen that Energy for female speaker is high as compared to male speaker. Likewise, Pitch of a female speaker is high as compared to male speaker.

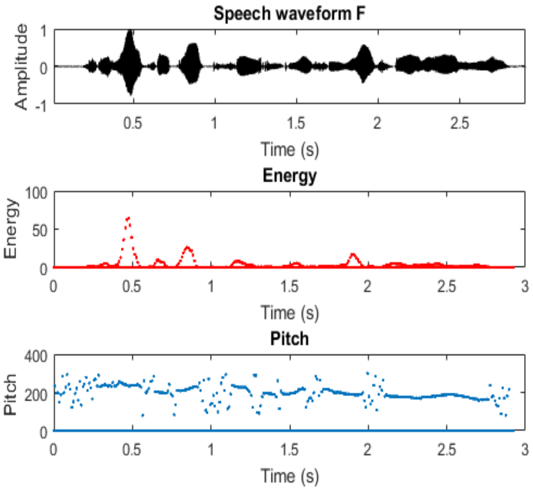

Fig. 5 (a)Speech signal of female voice (b) Energy plot (c)Pitch plot

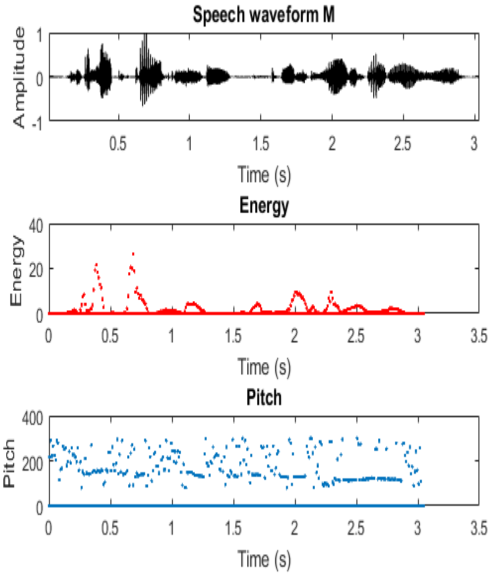

Fig. 6 (a) Speech signal of male voice (b) Energy plot (c) Pitch plot

B. Comparison of Mel Filter bank energies and Mel frequency cepstrum

Fig.7 and fig.8 shows the plot of Mel filter bank energy and Mel cepstrum for female and male speakers. On comparing Fig. 7 and Fig. 8,it can be seen that Mel filter bank energy of female is high as compared to male. Likewise, Mel frequency cepstrum of female is high as compared to a male speaker.

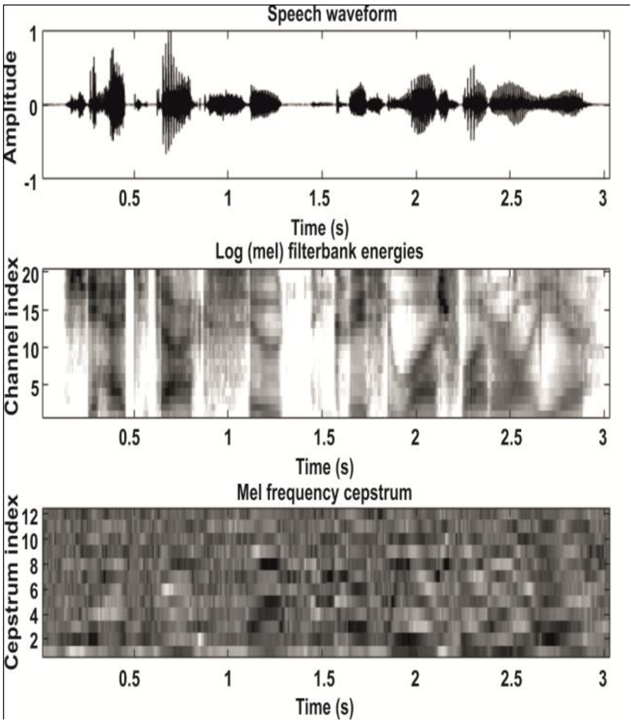

Fig. 7 (a) Speech signal of female voice (b) Mel filter bank energy plot (c) Mel frequency cepstrum plot.

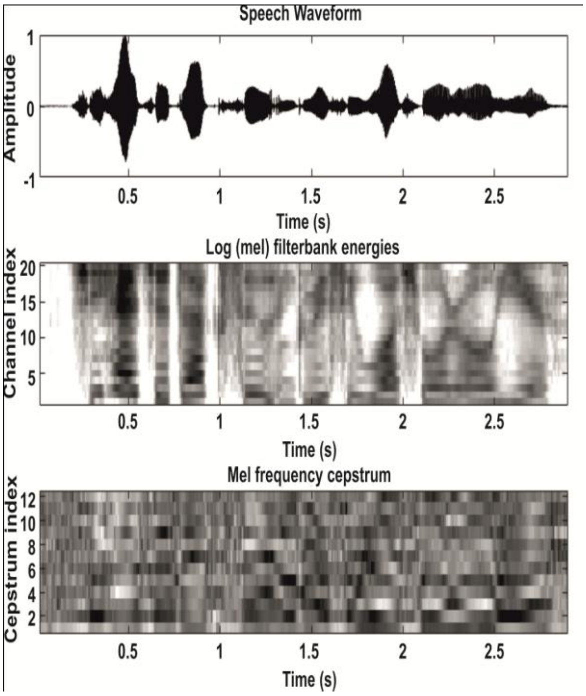

Fig.8 (a) Speech signal of male speaker (b) Mel filter bank energy plot (c) Mel frequency cepstrum plot.

VII. CONCLUSION

This work discusses about the existing features and classifiers used for the identification of gender of the speakers. For this, TIMIT database is used for all the training and testing files. A total of 280 files are used, outof which 140 are males 140 are females. It is found that using speech signal features such as energy, pitch, mfcc, the accuracy upto 96.25% is achieved. This accuracy is achieved when training and testing is carried out on 80% and 20% of speech files respectively. More features and data can be used and analysed to get better accuracy.

VIII. FUTURE SCOPE

This work can be further extended by adding gender features in emotion recognition. Even under normal conditions, gender identification will help in finding the emotion of male and female in different situations. Further gender identification can be improved by adding more features and using classifier like Gaussian mixture model and Hidden markov model along with SVM classifier to enhance the accuracy further.

REFERENCES

[1] E.S.Paris, and M.J.Carey, "Language independent gender identification'', International Conference on Acoustics Speech and Signal Processing Conference Proceedings, vol. 2, pp. 685-688,1996.

[2] S.Slomka and S.Sridharan, "Automatic gender

identification optimised for language independence'', IEEE TENCON Speech and Image Technologies for Computing and Telecommunications, pp. 145-148, 1997.

[3] R.D.R.Fayunder, A.A.C.Martins, "Automatic gender identification by speech signal using eigen filtering based on hebbian learning'', Proceedings of the 7th Brazilian Symposium on Neural Networks, pp. 212-216, 2002.

[4] H.Harb, and L.Chen, "Gender identification using a general audio classifier'', Proceedings of International conference on Multimedia and Expo, vol. 2, pp. 733- 736,2003.

[5] A.Lindgren, M.Johnson, and R.Povinelli, "Speech Recognition using reconstructed phase space features'',Speaker and Signal Processing Proceedings,
vol. 1, pp. 60-63,2003.

[6] H.Harb, and C.Chen, "Voiced based gender identification in multimedia applications'', Journal of Intelligent Information Systems, pp. 179-198, 2005.

[7] F.Yingle, Y.Li, and T.Qinge, "Speaker Gender Identification based on combining linear and non linear
features'', Proceedings of the 7th World Congress on Intelligent Control and Automation, pp. 6745-6749, 2008.

[8] M.Abdollhi, E.Valavi, H.A.Noubari, "Voiced based gender identification via multiresolution frame classification of spectro temporal maps'', Proceedings of International Joint conference on Neural Networks, pp. 1-4, 2009.

[9]M.A.Keyvanarad, M.M.Homayoupous, "Improvement on automatic speaker gender identification using classifier fusion'', Proceedings of Iranian Conference on Electrical Engineering. pp. 538-541, 2010.

[10] R.Djemili, H.Bourouba, M.Cherif, A.Korba, "A speech signal based gender identification system using four classifier'', International Conference on Multimedia Computing and systems, pp.184-187, 2012.

[11] Ghosal.A, and Dutta.S, "Automatic male female voice discrimination", International conference on Issues and Challenges in Intelligent Computing Techniques, pp.731-735,2014.

[12] G.S.Archana, and M.Malleswari, "Gender identification and performance analysis of speech signals'',Global Conference on Communication Technologies, pp. 483-489, 2015.

[13] Kumari.M, and Talukdar.N, " A new gender detection algorithm considering the non-stationarity of speech signal", IEEE 2nd International conference on Comunication, Control and Intelligent Systems, pp. 141-146, 2016.

[14] Ramdinmawii.E, and Mittal.V.K, "Gender identification from speech signal by examining the speech production characteristics", Internattional conference on signal processing and communication, pp. 244-249, 2016.

[15] Shan.L, Longting.X and Zhen.Y, "Multidimensional Speaker Information Recognition based onProposed Baseline System'', IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference, pp. 1776-1780, 2017.